

# Compressed sensing quantum process tomography for superconducting quantum gates

Andrey V. Rodionov<sup>1</sup>, Andrzej Veitia<sup>1</sup>, R. Barends<sup>2</sup>, J. Kelly<sup>2</sup>, Daniel Sank<sup>2</sup>,

J. Wenner<sup>2</sup>, John M. Martinis<sup>2</sup>, Robert L. Kosut<sup>3</sup>, and Alexander N. Korotkov<sup>1</sup>

<sup>1</sup>*Department of Electrical Engineering, University of California, Riverside, California 92521, USA*

<sup>2</sup>*Department of Physics, University of California, Santa Barbara, California 93106, USA*

<sup>3</sup>*SC Solutions, 1261 Oakmead Parkway, Sunnyvale, California 94085, USA*

(Dated: July 4, 2014)

We apply the method of compressed sensing (CS) quantum process tomography (QPT) to characterize quantum gates based on superconducting Xmon and phase qubits. Using experimental data for a two-qubit controlled-Z gate, we obtain an estimate for the process matrix  $\chi$  with reasonably high fidelity compared to full QPT, but using a significantly reduced set of initial states and measurement configurations. We show that the CS method still works when the amount of used data is so small that the standard QPT would have an underdetermined system of equations. We also apply the CS method to the analysis of the three-qubit Toffoli gate with numerically added noise, and similarly show that the method works well for a substantially reduced set of data. For the CS calculations we use two different bases in which the process matrix  $\chi$  is approximately sparse, and show that the resulting estimates of the process matrices match each other with reasonably high fidelity. For both two-qubit and three-qubit gates, we characterize the quantum process by not only its process matrix and fidelity, but also by the corresponding standard deviation, defined via variation of the state fidelity for different initial states.

PACS numbers: 03.65.Wj, 03.67.Lx, 85.25.Cp

## I. INTRODUCTION

An important challenge in quantum information science and quantum computing is the experimental realization of high-fidelity quantum operations on multi-qubit systems. Quantum process tomography (QPT) [1–3] is a procedure devised to fully characterize a quantum operation. The role of QPT in experimental characterization of quantum gates is twofold. First, it allows us to quantify the quality of the gate; that is, it tells us how close the actual and desired quantum operations are. Second, QPT may aid in diagnosing and correcting errors in the experimental operation [4–8]. The importance of QPT has led to extensive theoretical research on this subject (e.g., [9–14]).

Although conceptually simple, QPT suffers from a fundamental drawback: the number of required experimental configurations scales exponentially with the number of qubits (e.g., [15]). An  $N$ -qubit quantum operation can be represented by a  $4^N \times 4^N$  process matrix  $\chi$  [1] containing  $16^N$  independent real parameters (or  $16^N - 4^N$  parameters for a trace-preserving operation) which can be determined experimentally by QPT. Therefore, even for few-qubit systems, QPT involves collecting large amounts of tomographic data and heavy classical postprocessing. To alleviate the problem of exponential scaling of QPT resources, alternative methods have been developed, e.g., randomized benchmarking [16–18] and Monte Carlo process certification [19, 20]. These protocols, however, find only the fidelity of an operation instead of its full process matrix. Both randomized benchmarking and Monte Carlo process certification have been demonstrated experimentally for superconducting qubit gates (see [21–23] and references therein). Although these protocols are efficient tools for the verification of quantum gates, their limitation lies in the fact that they do not provide any description of particular errors affecting a given process and therefore they cannot

be used to improve the performance of the gates.

Recently, a new approach to QPT which incorporates ideas from signal processing theory has been proposed [24, 25]. The basic idea is to combine standard QPT with compressed sensing (CS) theory [26–29], which asserts that sparse signals may be efficiently recovered even when heavily under-sampled. As a result, compressed sensing quantum process tomography (CS QPT) enables one to recover the process matrix  $\chi$  from far fewer experimental configurations than standard QPT. The method proposed in [24, 25] is hoped to provide an exponential speed-up over standard QPT. In particular, for a  $d$ -dimensional system the method is supposed to require only  $O(s \log d)$  experimental probabilities to produce a good estimate of the process matrix  $\chi$ , if  $\chi$  is known to be  $s$ -compressible [30] in some known basis. (For comparison, standard QPT requires at least  $d^4$  probabilities, where  $d = 2^N$  for  $N$  qubits.) Note that there are bases in which the process matrix describing the target process (the desired unitary operation) is maximally sparse, i.e. containing only one non-zero element; for example, this is the case for the so-called singular-value-decomposition (SVD) basis [24] and the Pauli-error basis [8]. Therefore, if the actual process is close to the ideal (target) process, then it is plausible to expect that its process matrix is approximately sparse when written in such a basis [25]. The CS QPT method was experimentally validated in Ref. [25] for a photonic two-qubit controlled-Z (CZ) gate. In that experiment, sufficiently accurate estimates for the process matrix were obtained via CS QPT using much fewer experimental configurations than the standard QPT.

The CS idea also inspired another (quite different) algorithm for quantum state tomography (QST) [31, 32], which can be generalized to QPT [32, 33]. This matrix-completion method of CS QST estimates the density matrices of nearly pure (low rank  $r$ )  $d$ -dimensional quantum states from expectation values of only  $O(rd \text{ poly } \log d)$  observables, instead of  $d^2$  observables required for standard QST. It is important

to mention that this method does not require any assumption about the quantum state of a system, except that it must be a low-rank state (in particular, we do not need to know the state approximately). The CS QST method has been used to reconstruct the quantum states of a 4-qubit photonic system [34] and cesium atomic spins [35]. In Ref. [32] it has been shown that using the Jamiołkowski process-state isomorphism [36] the formalism of CS QST can also be applied to the QPT, requiring  $O(rd^2 \text{ poly } \log d)$  measured probabilities (where  $r$  is the rank of the Jamiołkowski state) to produce a good estimate of the process matrix  $\chi$ . Therefore there is crudely a square-root speedup compared with standard QPT. Note that this algorithm requires exponentially more resources than the CS QPT method of Ref. [25], but it does not require to know a particular basis in which the matrix  $\chi$  is sparse. The performance of these two methods has been compared in the recent paper [33] for a simulated quantum system with dimension  $d = 5$ ; the reported result is that the method of Ref. [32] works better for coherent errors, while the method of Ref. [25] is better for incoherent errors.

In this paper we apply the method of Ref. [25] to the two-qubit CZ gate realized with superconducting qubits. Using the experimental results, we find that CS QPT works reasonably well when the number of used experimental configurations is up to  $\sim 7$  times less than for standard QPT. Using simulations for a three-qubit Toffoli gate, we find that the reduction factor is  $\sim 40$ , compared with standard QPT. In the analysis we calculate two fidelities: the fidelity of the CS QPT-estimated process matrix  $\chi_{\text{CS}}$  compared with the matrix  $\chi_{\text{full}}$  from the full data set and compared with  $\chi_{\text{ideal}}$  for the ideal unitary process. Besides calculating the fidelities, we also calculate the standard deviation of the fidelity, defined via the variation of the state fidelity for different initial states. We show that this characteristic is also estimated reasonably well by using the CS QPT.

Our paper is structured as follows. Section II is a brief review of standard QPT and CS QPT. In Sec. III we discuss the set of measurement configurations used to collect QPT data for superconducting qubits, and also briefly discuss our way to compute the process matrix  $\chi$  via compressed sensing. In Sec. IV we present our numerical results for the CS QPT of a superconducting two-qubit CZ gate. In this section we also compare numerical results obtained by applying the CS QPT method in two different operator bases, the Pauli-error basis and the SVD basis. In Sec. V we study the CS QPT of a simulated three-qubit Toffoli gate with numerically added noise. Then in Sec. VI we use the process matrices obtained via compressed sensing to estimate the standard deviation of the state fidelity, with varying initial state. Section VII is a conclusion. In Appendices we discuss the Pauli-error basis (Appendix A), SVD basis (Appendix B), and calculation of the average square of the state fidelity (Appendix C).

## II. METHODS OF QUANTUM PROCESS TOMOGRAPHY

### A. Standard Quantum Process Tomography

The idea behind QPT is to reconstruct a quantum operation  $\rho^{\text{in}} \mapsto \rho^{\text{fin}} = \mathcal{E}(\rho^{\text{in}})$  from experimental data. The quantum operation is a completely positive map, which for an  $N$ -qubit system prepared in the state with density matrix  $\rho^{\text{in}}$  can be written as

$$\mathcal{E}(\rho^{\text{in}}) = \sum_{\alpha, \beta=1}^{d^2} \chi_{\alpha\beta} E_{\alpha} \rho^{\text{in}} E_{\beta}^{\dagger}, \quad (1)$$

where  $d = 2^N$  is the dimension of the system,  $\chi \in \mathbb{C}^{d^2 \times d^2}$  is the process matrix and  $\{E_{\alpha} \in \mathbb{C}^{d \times d}\}$  is a chosen basis of operators. We assume that this basis is orthogonal,  $\langle E_{\alpha} | E_{\beta} \rangle \equiv \text{Tr}(E_{\alpha}^{\dagger} E_{\beta}) = Q \delta_{\alpha\beta}$ , where  $Q = d$  for the Pauli basis and Pauli-error basis, while  $Q = 1$  for the SVD basis (see Appendices A and B). Note that for a trace-preserving operation  $\text{Tr}(\chi) = 1$  if  $Q = d$ , while  $\text{Tr}(\chi) = d$  if  $Q = 1$ . In this paper we implicitly assume the usual normalization  $Q = d$ , unless mentioned otherwise. The process matrix  $\chi$  is positive semidefinite (which implies being Hermitian), and we also assume it to be trace-preserving,

$$\chi \geq 0 \quad (\text{positive semidefinite}), \quad (2)$$

$$\sum_{\alpha, \beta=1}^{d^2} \chi_{\alpha\beta} E_{\beta}^{\dagger} E_{\alpha} = \mathbb{I}_d \quad (\text{trace preserving}). \quad (3)$$

These conditions ensure that  $\rho^{\text{fin}} = \mathcal{E}(\rho^{\text{in}})$  is a legitimate density matrix for an arbitrary (legitimate) input state  $\rho^{\text{in}}$ . The condition (3) reduces the number of real independent parameters in  $\chi$  from  $d^4$  to  $d^4 - d^2$ . Hence, the number of parameters needed to fully specify the map  $\mathcal{E}$  scales as  $O(16^N)$  with the number of qubits  $N$ . Note that the set of allowed process matrices  $\chi$  defined by Eqs. (2) and (3) is convex [24, 37].

The essential idea of standard QPT is to exploit the linearity of the map (1) by preparing the qubits in different initial states, applying the quantum gate, and then measuring a set of observables until the collected data allows us to obtain the process matrix  $\chi$  through matrix inversion or other methods. More precisely, if the qubits are prepared in the state  $\rho_k^{\text{in}}$ , then the probability of finding them in the (measured) state  $|\phi_i\rangle$  after applying the gate is given by

$$P_{ik} = \text{Tr}(\Pi_i \mathcal{E}(\rho_k^{\text{in}})) = \sum_{\alpha, \beta} \text{Tr}(\Pi_i E_{\alpha} \rho_k^{\text{in}} E_{\beta}^{\dagger}) \chi_{\alpha\beta}, \quad (4)$$

where  $\Pi_i = |\phi_i\rangle \langle \phi_i|$ . By preparing the qubits in one of the linearly independent input states  $\{\rho_1^{\text{in}}, \dots, \rho_{N_{\text{in}}}^{\text{in}}\}$  and performing a series of projective measurements  $\{\Pi_1, \dots, \Pi_{N_{\text{meas}}}\}$  on the output states, one obtains a set of  $m = N_{\text{in}} N_{\text{meas}}$  probabilities  $\{P_{ik}\}$  which, using Eq. (4), may be written as

$$\vec{P}(\chi) = \Phi \vec{\chi}, \quad (5)$$

where  $\vec{P}(\chi) \in \mathbb{C}^{m \times 1}$  and  $\vec{\chi} \in \mathbb{C}^{d^4 \times 1}$  are vectorized forms of  $\{P_{ik}\}$  and  $\chi_{\alpha\beta}$ , respectively. The  $m \times d^4$  transformation matrix  $\Phi$  has entries given by  $\Phi_{ik,\alpha\beta} = \text{Tr}(\Pi_i E_\alpha \rho_k^{\text{in}} E_\beta^\dagger)$ .

In principle, for tomographically complete sets of input states  $\{\rho_1^{\text{in}}, \dots, \rho_{N_{\text{in}}}^{\text{in}}\}$  and measurement operators  $\{\Pi_1, \dots, \Pi_{N_{\text{meas}}}\}$ , one could invert Eq. (5) and thus uniquely find  $\chi$  by using the experimental set of probabilities  $\vec{P}^{\text{exp}}$ . In practice, however, because of experimental uncertainties present in  $\vec{P}^{\text{exp}}$ , the process matrix thus obtained may be non-physical, that is, inconsistent with the conditions (2) and (3). In standard QPT this problem is remedied by finding the *physical* process matrix [satisfying (2) and (3)] that minimizes (in some sense) the difference between the probabilities  $\vec{P}(\chi)$  and the experimental probabilities  $\vec{P}^{\text{exp}}$ .

Two popular methods used to estimate a physical process matrix  $\chi$  compatible with the experimental data are the maximum likelihood (ML) method [38–40] (see also [41, 42]) and the least-squares (LS) method [21, 43, 44]. The ML method minimizes the cost function [38]

$$\mathcal{C}_{ML} = -\sum_j P_j^{\text{exp}} \ln P_j(\chi), \quad (6)$$

where the index  $j$  labels the measured probabilities, while the LS method (often also called maximum likelihood) minimizes the difference between  $\vec{P}(\chi)$  and  $\vec{P}^{\text{exp}}$  in the  $\ell_2$ -norm sense [45], so the minimized cost function is

$$\mathcal{C}_{LS} = \|\vec{P}(\chi) - \vec{P}^{\text{exp}}\|_{\ell_2}^2 = \sum_j [P_j^{\text{exp}} - P_j(\chi)]^2. \quad (7)$$

In both methods the conditions (2) and (3) should be satisfied to ensure that  $\chi$  corresponds to a physical process. This can be done in a number of ways, for example, using the Cholesky decomposition, or Lagrange multipliers, or just stating the conditions (2) and (3) as a constraint (if an appropriate software package is used). The ML method (6) is natural when the inaccuracy of  $\vec{P}^{\text{exp}}$  is dominated by the statistical error due to a limited number of experimental runs. However, this method does not work well if a target probability  $P_j$  is near zero, but  $P_j^{\text{exp}}$  is not near zero due to experimental imperfections (e.g., “dark counts”); this is because the cost function (6) is very sensitive to changes in  $P_j^{\text{exp}}$  when  $P_j(\chi) \approx 0$ . Therefore, the LS method (7) is a better choice when the inaccuracy of  $\vec{P}^{\text{exp}}$  is not due to a limited number of experimental runs.

Note that other cost functions can also be used for minimization in the procedure. For example, by replacing  $\ln P_j(\chi)$  in Eq. (6) with  $\ln[P_j(\chi)/P_j^{\text{exp}}]$  (this obviously does not affect optimization), then expanding the logarithm to second order, and using condition  $\sum_j P_j(\chi) = \sum_j P_j^{\text{exp}}$  (which cancels the first-order term), we obtain [41]  $\mathcal{C}_{ML} \approx \text{const} + \sum_j [P_j^{\text{exp}} - P_j(\chi)]^2 / 2P_j^{\text{exp}}$ . This leads to another natural cost function

$$\mathcal{C} = \sum_j \frac{[P_j(\chi) - P_j^{\text{exp}}]^2}{P_j^{\text{exp}} + a}, \quad (8)$$

where we phenomenologically introduced an additional parameter  $a$ , so that for  $a \gg 1$  the minimization reduces to the LS method, while for  $a \ll 1$  it is close to the ML method (the parameter  $a$  characterizes the relative importance of non-statistical and statistical errors). One more natural cost function is similar to Eq. (8), but with  $P_j^{\text{exp}}$  in the denominator replaced by  $P_j^{\text{exp}}(1 - P_j^{\text{exp}})$  (see [38]), which corresponds to the binomial distribution variance.

In this paper we use the LS method (7) for the standard QPT. In particular, we find the process matrix  $\chi_{\text{full}}$  for the full data set  $\vec{P}_{\text{full}}^{\text{exp}}$  by minimizing  $\|\vec{P}(\chi_{\text{full}}) - \vec{P}_{\text{full}}^{\text{exp}}\|_{\ell_2}$ , subject to conditions Eqs. (2) and (3). Note that such minimization is a convex optimization problem and therefore computationally tractable.

## B. Compressed Sensing Quantum Process Tomography

If the number of available experimental probabilities is smaller than the number of independent parameters in the process matrix (i.e.  $m < d^4 - d^2$ ), then the set of linear equations Eq. (5) for the process matrix  $\chi$  becomes underdetermined. Actually, the LS method may still formally work in this case for some range of  $m$ , but, as discussed in Secs. IV E and V, it is less effective.

By using the ideas of compressed sensing [26–29], the method of CS QPT requires a significantly smaller set of experimental data to produce a reasonably accurate estimate of the process matrix. Let us formulate the problem mathematically as follows: we wish to find the physical process matrix  $\chi_0$  satisfying the equation

$$\vec{P}^{\text{exp}} = \Phi \vec{\chi}_0 + \vec{z}, \quad (9)$$

where the vector  $\vec{P}^{\text{exp}} \in \mathbb{C}^m$  (with  $m < d^4 - d^2$ ) and the matrix  $\Phi \in \mathbb{C}^{m \times d^4}$  are given, while  $\vec{z} \in \mathbb{C}^m$  is an unknown noise vector, whose elements are assumed to be bounded (in the root-mean-square sense) by a known limit  $\varepsilon$ ,  $\|\vec{z}\|_{\ell_2} / \sqrt{m} \leq \varepsilon$ . While this problem seems to be ill-posed since the available information is both noisy and incomplete, in Ref. [26] it was shown that if the vector  $\chi_0$  is sufficiently sparse and the matrix  $\Phi$  satisfies the restricted isometry property (RIP),  $\chi_0$  can be accurately estimated from Eq. (9). Note that the CS techniques of Ref. [26] were developed in the context of signal processing; to adapt [24] these techniques to QPT we also need to include the positivity and trace-preservation conditions, Eqs. (2) and (3).

The idea of CS QPT [25] is to minimize the  $\ell_1$ -norm [45] of  $\vec{\chi}$  in a basis where  $\chi$  is assumed to be approximately sparse. Mathematically, the method is solving the following convex optimization problem:

$$\text{minimize } \|\vec{\chi}\|_{\ell_1}, \quad (10)$$

$$\text{subject to } \|\vec{P}(\chi) - \vec{P}^{\text{exp}}\|_{\ell_2} / \sqrt{m} \leq \varepsilon \quad (11)$$

and conditions (2), (3).

As shown in Refs. [25, 28], a faithful reconstruction recovery of an approximately  $s$ -sparse process matrix  $\chi_0$  via this optimization is guaranteed (see below) if (i) the matrix  $\Phi$  satisfies



the RIP condition,

$$1 - \delta_s \leq \frac{\|\Phi \vec{\chi}_1 - \Phi \vec{\chi}_2\|_{\ell_2}^2}{\|\vec{\chi}_1 - \vec{\chi}_2\|_{\ell_2}^2} \leq 1 + \delta_s, \quad (12)$$

for all  $s$ -sparse vectors (process matrices)  $\vec{\chi}_1$  and  $\vec{\chi}_2$ , (ii) the isometry constant  $\delta_s$  is sufficiently small,  $\delta_s < \sqrt{2} - 1$ , and (iii) the number of data points is sufficiently large,

$$m \geq C_0 s \log(d^4/s) = O(sN), \quad (13)$$

where  $C_0$  is a constant. Quantitatively, if  $\chi_{\text{CS}}$  is the solution of the optimization problem [Eqs. (10) and (11)], then the estimation error  $\|\chi_{\text{CS}} - \chi_0\|_{\ell_2}$  is bounded as

$$\frac{\|\chi_{\text{CS}} - \chi_0\|_{\ell_2}}{\sqrt{m}} \leq \frac{C_1 \|\chi_0(s) - \chi_0\|_{\ell_1}}{\sqrt{ms}} + C_2 \varepsilon, \quad (14)$$

where  $\chi_0(s)$  is the best  $s$ -sparse approximation of  $\chi_0$ , while  $C_1$  and  $C_2$  are constants of the order  $O(\delta_s)$ . Note that in the noiseless case ( $\varepsilon = 0$ ) the recovery is exact if the process matrix  $\chi_0$  is  $s$ -sparse. Also note that while the required number of data points  $m$  and the recovery accuracy depend on the sparsity  $s$ , the method itself [Eqs. (10) and (11)] does not depend on  $s$ , and therefore  $s$  need not be known.

The inequality (13) and the first term in the inequality (14) indicate that the CS QPT method is supposed to work well only if the actual process matrix  $\chi_0$  is sufficiently sparse. Therefore, it is important to use an operator basis  $\{E_\alpha\}$  [see Eq. (1)], in which the ideal (desired) process matrix  $\chi_{\text{ideal}}$  is maximally sparse, i.e., it contains only one nonzero element. Then it is plausible to expect the actual process matrix  $\chi_0$  to be approximately sparse [25]. In this paper we will use two bases in which the ideal process matrix is maximally sparse. These are the so-called Pauli-error basis [8] and the SVD basis of the ideal unitary operation [24]. In the Pauli-error basis  $\{E_\alpha\}$ , the first element  $E_1$  coincides with the desired unitary  $U$ , while other elements are related via the  $N$ -qubit Pauli matrices  $\mathcal{P}$ , so that  $E_\alpha = U\mathcal{P}_\alpha$ . In the SVD basis  $E_1 = U/\sqrt{d}$ , and other elements are obtained via a numerical SVD procedure. More details about the Pauli-error and SVD bases are discussed in Appendices A and B.

As mentioned previously, the method of CS QPT involves the RIP condition (12) for the transformation matrix  $\Phi$ . In Ref. [25] it was shown that if the transformation matrix  $\Phi$  in Eq. (5) is constructed from randomly selected input states  $\rho_k^{\text{in}}$  and random measurements  $\Pi_i$ , then  $\Phi$  obeys the RIP condition with high probability. Notice that once a basis  $\{E_\alpha\}$  and a tomographically complete (or overcomplete) set  $\{\rho_k^{\text{in}}, \Pi_i\}$  have been chosen, the matrix  $\Phi_{\text{full}}$  corresponding to the full data set is fully defined, since it does not depend on the experimental outcomes. Therefore, the mentioned above result of Ref. [25] tells us that if we build a matrix  $\Phi_m$  by randomly selecting  $m$  rows from  $\Phi_{\text{full}}$ , then  $\Phi_m$  is very likely to satisfy the RIP condition. Hence, the submatrix  $\Phi_m \in \mathbb{C}^{m \times d^4}$ , together with the corresponding set of experimental outcomes  $\vec{P}^{\text{exp}} \in \mathbb{C}^m$  can be used to produce an estimate of the process matrix via the  $\ell_1$ -minimization procedure (10) and (11).

### III. STANDARD AND CS QPT OF MULTI-QUBIT SUPERCONDUCTING GATES

There are several different ways to perform standard QPT for an  $N$ -qubit quantum gate realized with superconducting qubits [46–53]. The differences are the following. First, it can be performed using either  $n_{\text{in}} = 4$  initial states for each qubit [47–50], e.g.,  $\{|0\rangle, |1\rangle, (|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle + i|1\rangle)/\sqrt{2}\}$ , or using  $n_{\text{in}} = 6$  initial states per qubit [51, 52],  $\{|0\rangle, |1\rangle, (|0\rangle \pm |1\rangle)/\sqrt{2}, (|0\rangle \pm i|1\rangle)/\sqrt{2}\}$ , so that the total number of initial states is  $N_{\text{in}} = n_{\text{in}}^N$ . (It is tomographically sufficient to use  $n_{\text{in}} = 4$ , but the set of 6 initial states is more symmetric, so it can reduce the effect of experimental imperfections.) Second, the final measurement of the qubits can be realized in the computational basis after one out of  $n_{\text{R}} = 3$  rotations per qubit [47, 49], e.g.,  $\mathcal{R}_{\text{meas}} = \{\mathbb{I}, R_y^{-\pi/2}, R_x^{\pi/2}\}$ , or  $n_{\text{R}} = 4$  rotations [21, 48, 52], e.g.,  $\mathcal{R}_{\text{meas}} = \{\mathbb{I}, R_y^{\pi/2}, R_x^{\pi/2}, R_z^{\pi/2}\}$ , or  $n_{\text{R}} = 6$  rotations [46, 51, 53], e.g.,  $\mathcal{R}_{\text{meas}} = \{\mathbb{I}, R_y^{\pi/2}, R_x^{\pm\pi/2}, R_z^{\pm\pi/2}\}$ . This gives  $N_{\text{R}} = n_{\text{R}}^N$  measurement “directions” in the Hilbert space. Third, it may be possible to measure the state of each qubit simultaneously [46, 47, 49], so that the probabilities for all  $2^N$  outcomes are measured, or it may be technically possible to measure the probability for only one state (say,  $|0\dots 0\rangle$ ) or a weighed sum of the probabilities [48, 50, 51]. Therefore, the number of measured probabilities for each configuration is either  $N_{\text{prob}} = 2^N$  (with  $2^N - 1$  independent probabilities, since their sum is equal 1) or  $N_{\text{prob}} = 1$ . Note that if  $N_{\text{prob}} = 2^N$ , then using  $n_{\text{R}} = 6$  rotations per qubit formally gives the same probabilities as for  $n_{\text{R}} = 3$ , and in an experiment this formal symmetry can be used to improve the accuracy of the results. In contrast, in the case when  $N_{\text{prob}} = 1$ , the use of  $n_{\text{R}} = 4$  or  $n_{\text{R}} = 6$  are natural for the complete tomography.

Thus, the number of measurement configurations (including input state and rotations) in standard QPT is  $M_{\text{conf}} = N_{\text{in}} N_{\text{R}} = n_{\text{in}}^N n_{\text{R}}^N$ , while the total number of probabilities in the data set is  $M = M_{\text{conf}} N_{\text{prob}}$ . This number of probabilities can be as large as  $M = 72^N$  for  $n_{\text{in}} = 6$ ,  $n_{\text{R}} = 6$ , and  $N_{\text{prob}} = 2^N$  (with  $72^N - 36^N$  independent probabilities). Since only  $16^N - 4^N$  independent probabilities are necessary for the standard QPT, a natural choice for a shorter experiment is  $n_{\text{in}} = 4$ ,  $n_{\text{R}} = 3$ , and  $N_{\text{prob}} = 2^N$ ; then  $M = 24^N$ , with  $24^N - 12^N$  independent probabilities. If  $N_{\text{prob}} = 1$  due to the limitations of the measurement technique, then the natural choices are  $n_{\text{in}} = 4$  and  $n_{\text{R}} = 4$ , giving  $M = 16^N$  or  $n_{\text{in}} = 4$  and  $n_{\text{R}} = 6$ , giving  $M = 24^N$ .

In this paper we focus on the case  $n_{\text{in}} = 4$ ,  $n_{\text{R}} = 3$ , and  $N_{\text{prob}} = 2^N$ . Then for a two-qubit quantum gate there are  $M_{\text{conf}} = 12^N = 144$  measurement configurations and  $M = 24^N = 576$  probabilities (432 of them independent). For a three-qubit gate there are  $M_{\text{conf}} = 1728$  configurations and  $M = 13824$  probabilities (12096 of them independent).

The main experimental data used in this paper are for the two-qubit CZ gate realized with Xmon qubits [54]. The data were obtained with  $n_{\text{in}} = 6$ ,  $n_{\text{R}} = 6$ , and  $N_{\text{prob}} = 2^N$ . However, since the main emphasis of this paper is analysis

of the QPT with a reduced data set, we started by reducing the data set to  $n_{\text{in}} = 4$  and  $n_{\text{R}} = 3$  by using only the corresponding probabilities and removing other data. We will refer to these data as “full data” (with  $M_{\text{conf}} = 144$  and  $M = 24^N = 576$ ). For testing the CS method we randomly choose  $m_{\text{conf}} \leq M_{\text{conf}}$  configurations, with corresponding  $m = 4m_{\text{conf}}$  experimental probabilities ( $3m_{\text{conf}}$  of them independent). Since the process matrix  $\chi$  is characterized by  $16^N - 4^N = 240$  independent parameters, the power of the CS method is most evident when  $m_{\text{conf}} < 80$ , so that the system of equations (5) is underdetermined. [For a three-qubit gate the system of equations becomes underdetermined for  $m_{\text{conf}} < (16^N - 4^N)/(2^N - 1) = 576$ .]

The data used for the analysis here were taken on a different device from the one used in Ref. [55]. For the device used here the qubits were coupled via a bus, and the entangling gate between qubits A and B was implemented with three multiqubit operations: 1) swap state from qubit B to bus, 2) CZ gate between qubit A and bus, 3) swap back to qubit B. The swap was done with the resonant Strauch gate [56], by detuning the frequency of qubit A with a square pulse. Generating a square pulse is experimentally challenging, moreover this gate has a single optimum in pulse amplitude and time. We also note that the qubit frequency control was not optimized for imperfections in the control wiring, as described in Ref. [57] (see also Fig. S4 in Supplementary Information of [55]). The combination of device, non-optimal control, and multiple operations, leads to the experimental process fidelity  $F_{\chi} = 0.91$  of the CZ gate used for the analysis here to be significantly less than the randomized benchmarking fidelity  $F_{\text{RB}} = 0.994$  reported in [55]. Moreover, QPT necessarily includes state preparation and measurement (SPAM) errors [18], while randomized benchmarking does not suffer from these errors. This is why we intentionally used the data for a not-well-optimized CZ gate so that the gate error dominates over the SPAM errors. (Note that we use correction for the imperfect measurement fidelity [46]; however, it does not remove the measurement errors completely.) It should also be mentioned that in the ideal case  $1 - F_{\chi} = (1 - F_{\text{RB}}) \times (1 + 2^{-N})$ , so the QPT fidelity is supposed to be slightly less than the randomized benchmarking fidelity.

For the full data set, we first calculate the process matrix  $\chi_{\text{full}}$  by using the least-squares method described at the end of Sec. II A. For that we use three different operator bases  $\{E_{\alpha}\}$ : the Pauli basis, the Pauli-error basis, and the SVD basis. The pre-computed transformation matrix  $\Phi$  in Eq. (5) depends on the choice of the basis, thus giving a basis-dependent result for  $\chi_{\text{full}}$ . We then check that the results essentially coincide by converting  $\chi_{\text{full}}$  between the bases and calculating the fidelity between the corresponding matrices (the infidelity is found to be less than  $10^{-7}$ ). The fidelity between two process matrices  $\chi_1$  and  $\chi_2$  is defined as the square of the Uhlmann fidelity [58, 59],

$$F(\chi_1, \chi_2) = \left( \text{Tr} \sqrt{\chi_1^{1/2} \chi_2 \chi_1^{1/2}} \right)^2, \quad (15)$$

so that it reduces to  $F(\chi_1, \chi_2) = \text{Tr}(\chi_1 \chi_2)$  [60] when at least one of the process matrices corresponds to a unitary operation.

Since  $0 \leq F \leq 1$ , we refer to  $1 - F$  as the infidelity.

After calculating  $\chi_{\text{full}}$  for the full data set, we can calculate its fidelity compared to the process matrix  $\chi_{\text{ideal}}$  of the desired ideal unitary operation,  $F_{\chi} = F_{\text{full}} = F(\chi_{\text{full}}, \chi_{\text{ideal}})$ . This is the main number used to characterize the quality of the quantum operation.

Then we calculate the compressed-sensing process matrix  $\chi_{\text{CS}}$  by solving the  $\ell_1$ -minimization problem described by Eqs. (10) and (11), using the reduced data set. It is obtained from the full data set by randomly selecting  $m_{\text{conf}}$  configurations out of the full number  $M_{\text{conf}}$  configurations. We use the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  to quantify how well the process matrix  $\chi_{\text{CS}}$  approximates the matrix  $\chi_{\text{full}}$  obtained from full tomographic data. Additionally, we calculate the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  between  $\chi_{\text{CS}}$  and the ideal operation, to see how closely it estimates the process fidelity  $F_{\text{full}}$ , obtained using the full data set.

Since both the least-squares and the  $\ell_1$ -norm minimization are convex optimization problems [24, 61], they can be efficiently solved numerically. We used two ways for MATLAB-based numerical calculations: (1) using the package CVX [62], which calls the convex solver SeDuMi [63]; or (2) using the package YALMIP [64], which calls the convex solver SDPT3 [65]. In general, we have found that for our particular realization of computation, CVX with the solver SeDuMi works better than the combination YALMIP-SDPT3 (more details are below).

#### IV. RESULTS FOR TWO-QUBIT CZ GATE

In this section we present results for the experimental CZ gate realized with superconducting Xmon qubits [54, 55]. As explained above, the full data set consists of  $M = 576$  measured probabilities (432 of them independent), corresponding to  $M_{\text{conf}} = 4^2 \times 3^2 = 144$  configurations, with 4 probabilities (3 of them independent) for each configurations. The LS method using the full data set produces the process matrix  $\chi_{\text{full}}$ , which has the process fidelity  $F(\chi_{\text{full}}, \chi_{\text{ideal}}) = 0.907$  relative to the ideal CZ operation. Note that our full data set is actually a subset of an even larger data set (as explained in the previous section), and the  $\chi$  matrix calculated from the initial set corresponds to the process fidelity of 0.928; the difference gives a crude estimate of the overall accuracy of the procedure.

The CS method calculations were mainly done in the Pauli-error basis, using the CVX-SeDuMi combination for the  $\ell_1$ -norm minimization. This is what is implicitly assumed in this section, unless specified otherwise. Note that the CS-method optimization is very different from the LS method. Therefore, even for the full data set we would expect the process matrix  $\chi_{\text{CS}}$  to be different from  $\chi_{\text{full}}$ . Moreover,  $\chi_{\text{CS}}$  depends on the noise parameter  $\varepsilon$  [see Eq. (11)], which to some extent is arbitrary. To clarify the role of the parameter  $\varepsilon$ , we will first discuss the CS method applied to the full data set, with varying  $\varepsilon$ , and then discuss the CS QPT for a reduced data set, using either near-optimal or non-optimal values of  $\varepsilon$ .

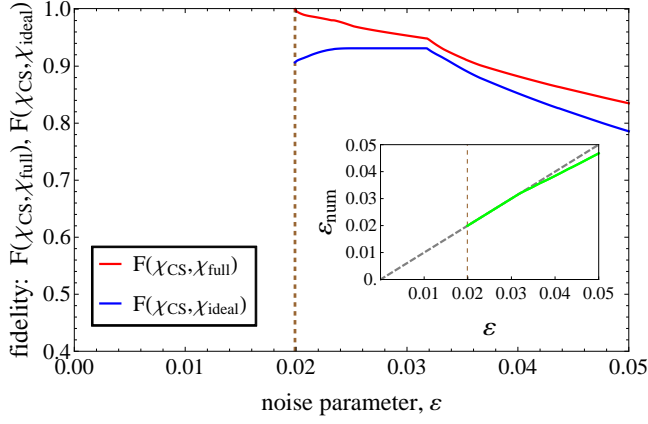


FIG. 1: (color online) The CS QPT procedure, applied to the full data set, with varying noise parameter  $\varepsilon$ . The red (upper) line shows the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  between the process matrix  $\chi_{\text{CS}}$  obtained using the compressed-sensing method and the matrix  $\chi_{\text{full}}$  obtained using the least-squares method. The blue (lower) line shows the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ , i.e., compared with the matrix  $\chi_{\text{ideal}}$  of the ideal unitary process. The vertical dashed brown line corresponds to the noise level  $\varepsilon_{\text{opt}} = \|\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{full}}\|_{\ell_2} / \sqrt{M} = 0.0199$  obtained in the LS procedure. The inset shows  $\varepsilon_{\text{num}} = \|\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{CS}}\|_{\ell_2} / \sqrt{M}$  as a function of  $\varepsilon$  (green line); for comparison, the dashed line shows the expected straight line,  $\varepsilon_{\text{num}} = \varepsilon$ . The numerical calculations have been carried out in the Pauli-error basis using CVX-SeDuMi package.

### A. Full data set, varying $\varepsilon$

We start with calculating the process matrix  $\chi_{\text{CS}}$  by solving the  $\ell_1$ -minimization problem, Eqs. (10) and (11), using the full data set and varying the noise parameter  $\varepsilon$ . The resulting matrix is compared with the LS result  $\chi_{\text{full}}$  and with the ideal matrix  $\chi_{\text{ideal}}$ . Figure 1 shows the corresponding fidelities  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  and  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  as functions of  $\varepsilon$ . We see that  $\chi_{\text{CS}}$  coincides with  $\chi_{\text{full}}$  [so that  $F(\chi_{\text{CS}}, \chi_{\text{full}}) = 1$ ] at the optimal value  $\varepsilon_{\text{opt}} = 0.0199$ . This is exactly the noise level corresponding to the LS procedure,  $\|\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{full}}\|_{\ell_2} / \sqrt{M} = 0.0199$ . With  $\varepsilon$  increasing above this level, the relative fidelity between  $\chi_{\text{CS}}$  and  $\chi_{\text{full}}$  decreases, but it remains above 0.95 for  $\varepsilon < 0.028$ . Correspondingly, the process fidelity reported by  $\chi_{\text{CS}}$ , i.e.  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ , also changes. It starts with  $F(\chi_{\text{CS}}, \chi_{\text{ideal}}) = F(\chi_{\text{full}}, \chi_{\text{ideal}}) = 0.907$  for  $\varepsilon = 0.0199$ , then increases with increasing  $\varepsilon$ , then remains flat above  $\varepsilon = 0.025$ , and then decreases at  $\varepsilon > 0.032$ . We note that for another set of experimental data (for a CZ gate realized with phase qubits) there was no increasing part of this curve, and the dependence of  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  on  $\varepsilon$  remained practically flat for a wide range of  $\varepsilon$ ; one more set of experimental data for phase qubits again had the increasing part of this curve.

To check how close the result of  $\ell_1$ -optimization (10) is to the upper bound of the condition (11), we calculate the numerical value  $\varepsilon_{\text{num}} = \|\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{CS}}\|_{\ell_2} / \sqrt{M}$  as a function of  $\varepsilon$ . The result is shown in the inset of Fig. 1, we see that  $\varepsilon_{\text{num}}$

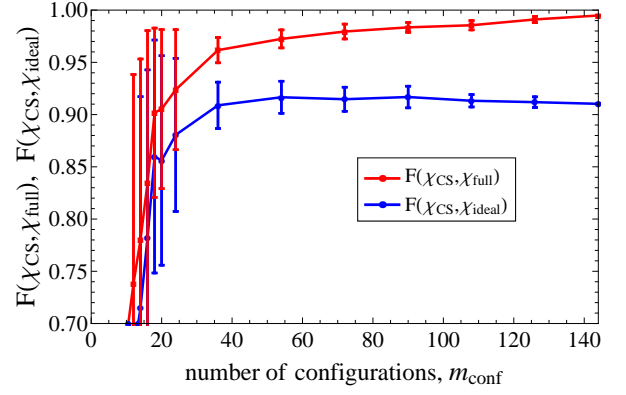


FIG. 2: (color online) The CS method results using a reduced data set with randomly chosen  $m_{\text{conf}}$  configurations. The red (upper) line shows the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  between the CS-estimated process matrix  $\chi_{\text{CS}}$  and the matrix  $\chi_{\text{full}}$  obtained from the full data set. The blue (lower) line shows the estimated process fidelity  $F_{\chi} = F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ . The procedure of randomly choosing  $m_{\text{conf}}$  out of 144 configurations is repeated 50 times; the error bars show the calculated standard deviations. The noise parameter  $\varepsilon = 0.02015$  is chosen slightly above  $\varepsilon_{\text{opt}}$ . The calculations are carried out in the Pauli-error basis using CVX-SeDuMi. The experimental data are for the CZ gate realized with Xmon qubits; the process fidelity is  $F(\chi_{\text{full}}, \chi_{\text{ideal}}) = 0.907$ .

is quite close to  $\varepsilon$ . The CVX-SeDuMi package does not solve the optimization problem for values of the noise parameter  $\varepsilon$  below the optimal value  $\varepsilon_{\text{opt}}$ .

Finding a proper value of  $\varepsilon$  to be used in the CS method is not a trivial problem, since for the reduced data set we cannot find  $\varepsilon_{\text{opt}}$  in the way we used. Therefore, the value of  $\varepsilon$  should be estimated either from some prior information about the noise level in the system or by trying to solve the  $\ell_1$ -minimization problem with varying value of  $\varepsilon$ . Note that the noise level  $\|\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{ideal}}\|_{\ell_2} / \sqrt{M}$  defined by the ideal process is not a good estimate of  $\varepsilon_{\text{opt}}$ ; in particular for our full data it is 0.035, which is significantly higher than  $\varepsilon_{\text{opt}} = 0.0199$ .

### B. Reduced data set, near-optimal $\varepsilon$

Now we apply the CS method to a reduced data set, by randomly choosing  $m_{\text{conf}}$  out of  $M_{\text{conf}} = 144$  configurations, while using all 4 probabilities for each configuration. (Therefore the number of used probabilities is  $m = 4m_{\text{conf}}$  instead of  $M = 4M_{\text{conf}}$  in the full data set.) For the noise level  $\varepsilon$  we use a value slightly larger than  $\varepsilon_{\text{opt}}$  [25]. If a value too close to  $\varepsilon_{\text{opt}}$  is used, then the optimization procedure often does not find a solution; this happens when we choose configurations with a relatively large level of noise in the measured probability values. For the figures presented in this subsection we used  $\varepsilon = 0.02015$ , which for the full data set corresponds to the fidelity of 0.995 compared with  $\chi_{\text{full}}$  and to the process fidelity of 0.910 (see Fig. 1).

Figure 2 shows the fidelities  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  (upper line) and



$F(\chi_{CS}, \chi_{ideal})$  (lower line) versus the number  $m_{conf}$  of used configurations. For each value of  $m_{conf}$  we repeat the procedure 50 times, choosing different random configurations. The error bars in Fig. 2 show the standard deviations ( $\pm\sigma$ ) calculated using these 50 numerical experiments, while the central points correspond to the average values.

We see that the upper (red) line starts with fidelity  $F(\chi_{CS}, \chi_{full}) = 0.995$  for the full data set ( $m_{conf} = 144$ ) and decreases with decreasing  $m_{conf}$ . It is important that this decrease is not very strong, so that we can reconstruct the process matrix reasonably accurately, using only a small fraction of the QPT data. We emphasize that the system of equations (5) in the standard QPT procedure becomes underdetermined at  $m_{conf} < 80$ ; nevertheless, the CS method reconstructs  $\chi_{full}$  quite well for  $m_{conf} \gtrsim 40$  and still gives reasonable results for  $m_{conf} \gtrsim 20$ . In particular, for  $m_{conf}$  between 40 and 80, the reconstruction fidelity  $F(\chi_{CS}, \chi_{full})$  changes between 0.96 and 0.98.

The lower (blue) line in Fig. 2 shows that the process fidelity  $F_\chi = F(\chi_{CS}, \chi_{ideal})$  can also be found quite accurately, using only  $m_{conf} \gtrsim 40$  configurations (the line remains practically flat), and the CS method still works reasonably well down to  $m_{conf} \gtrsim 20$ . Even though the blue line remains practically flat down to  $m_{conf} \simeq 40$ , the error bars grow, which means that in a particular experiment with substantially reduced set of QPT data, the estimated process fidelity  $F_\chi$  may noticeably differ from the actual value. It is interesting that the error bars become very large at approximately the same value ( $m_{conf} \simeq 20$ ), for which the average values for the red and blue lines become unacceptably low.

Figure 3 shows examples of the CS estimated process matrices  $\chi_{CS}$  for  $m_{conf} = 72$  (middle panel) and  $m_{conf} = 36$  (lower panel), together with the full-data process matrix  $\chi_{full}$  (upper panel). The process matrices are drawn in the Pauli-error basis to display the process imperfections more clearly. The peak  $\chi_{II,II}$  is off the scale and is cut arbitrarily. We see that the CS estimated process matrices are different from the full-data matrix; however the positions of the main peaks are reproduced exactly, and their heights are also reproduced rather well (for a small number of selected configurations the peaks sometimes appear at wrong positions). It is interesting to see that the CS procedure suppressed the height of minor peaks. Note that both presented  $\chi_{CS}$  are based on the data sets corresponding to underdetermined system of equations.

The computer resources needed for the calculation of results presented in Fig. 2 are not demanding. The calculations require about 30 MB of computer memory and 2–4 seconds time for a modest PC per individual calculation (smaller time for smaller number of configurations).

Besides the presented results, we have also performed analysis for the CS QPT of two CZ gates based on phase qubits. The results are qualitatively similar, except the process fidelity for phase-qubit gates was significantly lower: 0.62 and 0.51. The results for one of these gates are presented in Fig. 4. Comparing with Fig. 2, we see that CS QPT works better for this lower-fidelity gate. In particular, the blue line in Fig. 4 is practically flat down to  $m_{conf} \simeq 20$  and the error bars are quite small. We think that the CS QPT works better for a

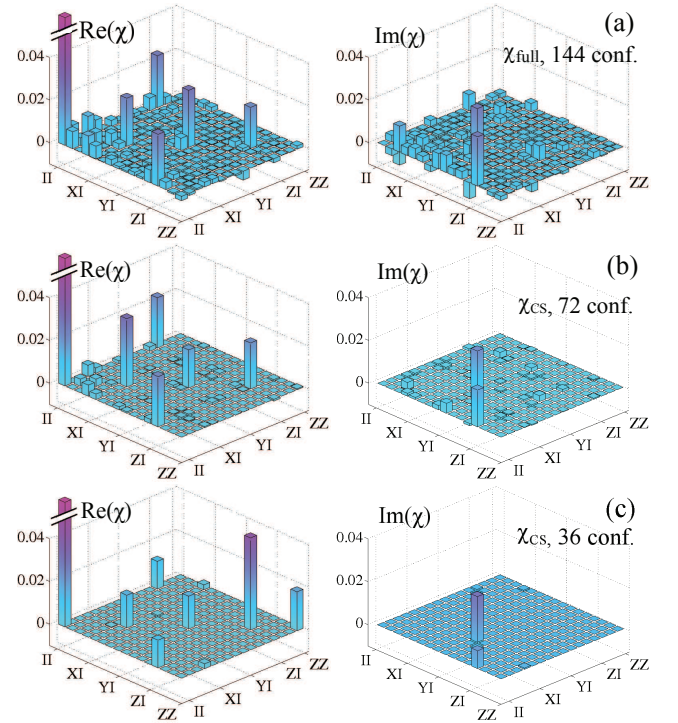


FIG. 3: (color online) (a) The process matrix  $\chi_{full}$  based on the full data set (144 configurations) and (b,c) the CS-estimated matrices  $\chi_{CS}$  using a reduced data set: 72 configurations (b) and 36 configurations (c). The process matrices are shown in the Pauli-error basis. The main element  $\chi_{II,II}$  (process fidelity) is off the scale and therefore is cut; its height is 0.907, 0.918, and 0.899 for the panels (a), (b), and (c), respectively. All other peaks characterize imperfections. The fidelity  $F(\chi_{CS}, \chi_{full})$  for the matrices in panels (b) and (c) is equal to 0.981 and 0.968, respectively. The middle and lower panels use the data set, corresponding to underdetermined systems of equations.

lower-fidelity gate because experimental imperfections affect the measurement error relatively less in this case than for a higher-fidelity gate.

Thus, our results show that for a CZ gate realized with superconducting qubits CS QPT can reduce the number of used QPT configurations by up to a factor of 7 compared with full QPT, and up to a factor of 4 compared with the threshold at which the system of equations for the standard QPT becomes underdetermined.

### C. Reduced data set, nonoptimal $\varepsilon$

As mentioned above, in a QPT experiment with a reduced data set, there is no straightforward way to find the near-optimal value of the noise parameter  $\varepsilon$  (which we find here from the full data set). Therefore, it is important to check how well the CS method works when a nonoptimal value of  $\varepsilon$  is used. Figure 5 shows the results similar to those in Fig. 2, but with several values of the noise parameter:  $\varepsilon/\varepsilon_{opt} = 1.01, 1.2, 1.4, 1.6, \text{ and } 1.8$ . The upper panel shows the fidelity between the matrix  $\chi_{CS}$  and the full-data matrix  $\chi_{full}$ ; the lower

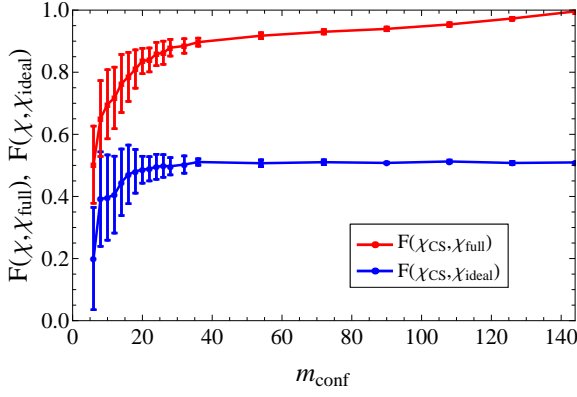


FIG. 4: (color online) Similar to Fig. 2, but for the CZ gate realized with superconducting phase qubits. The process fidelity  $F(\chi_{\text{full}}, \chi_{\text{ideal}}) = 0.51$  is much lower than that for the Xmon qubit gate. As we see, CS QPT works significantly better for this lower-fidelity gate than for the better gate presented in Fig. 2.

panel shows the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ . We see that the fidelity of the  $\chi$  matrix estimation,  $F(\chi_{\text{CS}}, \chi_{\text{full}})$ , becomes monotonously worse with increasing  $\varepsilon$ , while the estimated process fidelity,  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ , may become larger when a nonoptimal  $\varepsilon$  is used.

Similar results (not presented here) for the CZ gate based on phase qubits (see Fig. 4) have shown significantly better tolerance to a nonoptimal choice of  $\varepsilon$ ; in particular, even for  $\varepsilon = 3\varepsilon_{\text{opt}}$  the process fidelity practically coincides with the blue line in Fig. 4 (obtained for  $\varepsilon \approx \varepsilon_{\text{opt}}$ ). We believe the lower gate fidelity for phase qubits is responsible for this relative insensitivity to the choice of  $\varepsilon$ .

#### D. Comparison between Pauli-error and SVD bases

So far for the CS method we have used the Pauli-error basis, in which the process matrix  $\chi$  is expected to be approximately sparse because the ideal process matrix  $\chi_{\text{ideal}}$  contains only one non-zero element,  $\chi_{\text{ideal}, II, II} = 1$ . However, there are an infinite number of the operator bases with this property: for example, the SVD basis (see Appendix B) suggested in Refs. [24] and [25]. The process matrix is different in the Pauli-error and SVD bases, therefore the CS optimization should produce different results. To compare the results, we do the CS optimization in the SVD basis, then convert the resulting matrix  $\chi$  into the Pauli-error basis, and calculate the fidelity  $F(\chi_{\text{CS-SVD}}, \chi_{\text{CS}})$  between the transformed process matrix and the matrix  $\chi_{\text{CS}}$  obtained using optimization in the Pauli-error basis directly.

The green line in Fig. 6 shows  $F(\chi_{\text{CS-SVD}}, \chi_{\text{CS}})$  as a function of the selected size of the data set for the CZ gate realized with Xmon qubits, similar to Fig. 2 (the same  $\varepsilon$  is used). We also show the fidelity between the SVD-basis-obtained matrix  $\chi_{\text{CS-SVD}}$  and the full-data matrix  $\chi_{\text{full}}$  as well as the ideal process matrix  $\chi_{\text{ideal}}$ . For comparison we also include the lines shown in Fig. 2 (dashed lines), obtained using the Pauli-

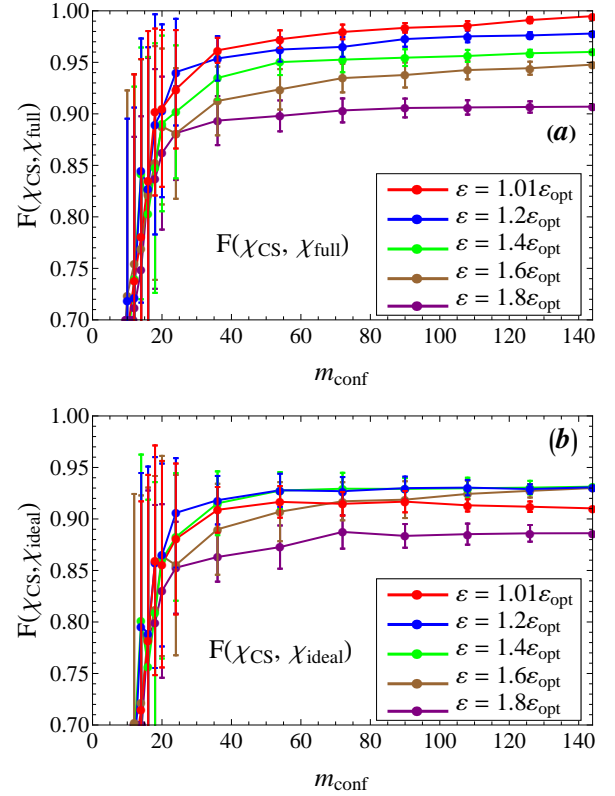


FIG. 5: (color online) (a) Fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  of the process matrix estimation and (b) the estimated process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  as functions of the data set size for several values of the noise parameter  $\varepsilon$  used in the CS optimization:  $\varepsilon/\varepsilon_{\text{opt}} = 1.01, 1.2, 1.4, 1.6$ , and  $1.8$ . Error bars show the standard deviations calculated using 50 random selections of reduced data sets. The red lines are the same as the lines in Fig. 2.

error basis. As we see, the results obtained in the two bases are close to each other, though the SVD basis seems to work a little better at small data sizes,  $m_{\text{conf}} \simeq 20$ . The visual comparison of  $\chi$ -matrices obtained in these bases (as in Fig. 3, not presented here) also shows that they are quite similar. It should be noted that the calculations in the SVD basis are somewhat faster ( $\sim 2$  seconds per point) and require less memory ( $\sim 6$  MB) than the calculations in the Pauli-error basis. This is because the matrix  $\Phi$  defined in Eq. (5) for the CZ gate contains about half the number of non-zero elements in the SVD basis than in the Pauli-error basis.

All results presented here are obtained using the CVX-SeaDuMi package. The results for the CZ gate obtained using the YALMIP-SDPT3 package are similar when the same value of  $\varepsilon$  is used. Surprisingly, in our realization of computation, the YALMIP-SDPT3 package still finds reasonable solutions when  $\varepsilon$  is significantly smaller than  $\varepsilon_{\text{opt}}$  (even when  $\varepsilon$  is zero or negative), so that the problem cannot have a solution; apparently in this case the solver increases the value of  $\varepsilon$  until a solution is found. This may seem to be a good feature of YALMIP-SDPT3. However, using  $\varepsilon < \varepsilon_{\text{opt}}$  should decrease the accuracy of the result (see the next subsection).



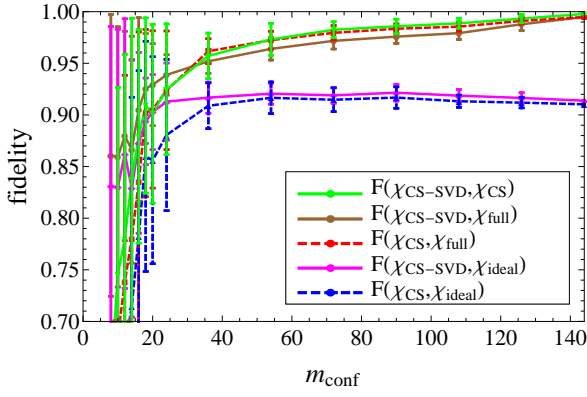


FIG. 6: (color online) Comparison between the CS results obtained in the SVD and Pauli-error bases. The green line shows the relative fidelity  $F(\chi_{\text{CS-SVD}}, \chi_{\text{CS}})$  as a function of the number  $m_{\text{conf}}$  of randomly selected configurations. We also show the fidelities  $F(\chi_{\text{CS-SVD}}, \chi_{\text{full}})$  (brown line),  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  (red dashed line), and process fidelities  $F(\chi_{\text{CS-SVD}}, \chi_{\text{ideal}})$  (magenta line) and  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  (blue dashed line). The dashed lines have been shown in Fig. 2. The results using the SVD basis are somewhat more accurate than those for the Pauli-error basis when  $m_{\text{conf}} < 40$ .

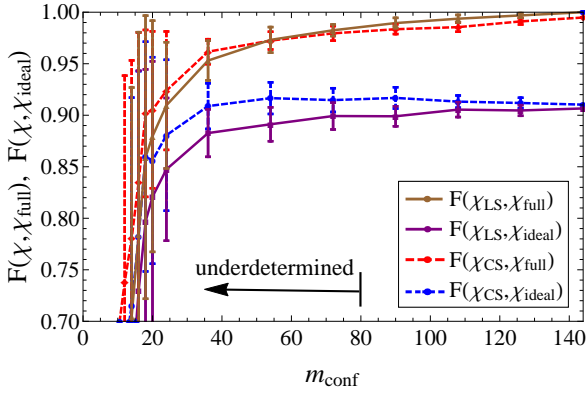


FIG. 7: (color online) Comparison between the results obtained by the LS and CS methods. The solid lines are for the LS method, the dashed lines (same as in Fig. 2) are for the CS method. The CS method is more accurate for a substantially reduced data set.

Moreover, YALMIP-SDPT3 does not work well for the Toffoli gate discussed in Sec. V. Thus we conclude that CVX-SeDuMi package is better than YALMIP-SDPT3 package for our CS calculations. (Note that this finding may be specific to our system.)

### E. Comparison with least-squares minimization

Besides using the CS method for reduced data sets, we also used the LS minimization [with constraints (2) and (3)] for the same reduced sets. Solid lines in Fig. 7 show the resulting fidelity  $F(\chi_{\text{LS}}, \chi_{\text{full}})$  compared with the full-data process matrix and the estimated process fidelity  $F(\chi_{\text{LS}}, \chi_{\text{ideal}})$ .

Somewhat surprisingly, the LS method still works (though

less well) in a significantly underdetermined regime. Naively, we would expect that in this case Eq. (5) can be satisfied exactly, and there are many exact solutions corresponding to the null space of the selected part of the matrix  $\Phi$ . However, numerical results show that in reality Eq. (5) cannot be satisfied exactly unless the selected data set is very small. The reason is that the matrix  $\chi$  has to be positive, and the (corrected) experimental probabilities can be close to the limits of the physical range or even outside it.

The problem is that the experimental probabilities are not directly obtained from the experiment, but are corrected for imperfect measurement fidelity [46]. As a result, they may become larger than one or smaller than zero. This happens fairly often for high fidelity gates because for an ideal operation the measurement results are often zeros and ones, so the experimental probabilities should also be close to zero or one. Any additional deviation due to imperfect correction for the measurement fidelity may then push the probabilities outside of the physical range. It is obvious that in this case Eq. (5) cannot be satisfied exactly for any physical  $\chi$ . To resolve this problem one could consider rescaling the probabilities in such instances, so that they are exactly one or zero instead of lying outside the range. However, this also does not help much because a probability of one means that the resulting state is pure, so this strongly reduces the number of free parameters in the process matrix  $\chi$ . As a result, Eq. (5) cannot be satisfied exactly, and the LS minimization is formally possible even in the underdetermined case.

Another reason why Eq. (5) may be impossible to satisfy in the underdetermined case, is that the randomly selected rows of the matrix  $\Phi$  can be linearly dependent. Then mathematically some linear relations between the experimental probabilities must be satisfied, while in reality they are obviously not satisfied exactly.

These reasons make the LS minimization a mathematically possible procedure even in the underdetermined regime. However, as we see from Fig. 7, in this case the procedure works less well than the compressed sensing, estimating the process matrix and process fidelity with a lower accuracy. Similar calculations for the CZ gate realized with phase qubits (not presented here) also show that the LS method does not work well at relatively small  $m_{\text{conf}}$ . The advantage of the compressed sensing in comparison with the LS minimization becomes even stronger for the three-qubit Toffoli gate considered in the next section. Note though that when the selected data set is large enough to give an overdetermined system of equations (5), the LS method works better than the CS method. Therefore, the compressed sensing is beneficial only for a substantially reduced (underdetermined) data set, which is exactly the desired regime of operation.

### V. THREE-QUBIT CS QPT FOR TOFFOLI GATE

In this section we apply the compressed sensing method to simulated tomographic data corresponding to a three-qubit Toffoli gate [1, 46, 50, 66, 67]. As discussed in Sec. III, the process matrix of a three-qubit gate contains  $16^3 - 4^3 = 4032$

independent real parameters, while the full QPT requires  $M_{\text{conf}} = 12^3 = 1728$  measurement configurations yielding a total of  $M = 12^3 \times 2^3 = 13824$  experimental probabilities, if we use  $n_{\text{in}} = 4$  initial states and  $n_{\text{R}} = 3$  measurement rotations per qubit, with all qubits measured independently. If we work with a partial data set, the system of equations (5) becomes underdetermined if the number  $m_{\text{conf}}$  of used configurations is less than  $4032/7 = 576$ . In such a regime the traditional maximum likelihood or LS methods are not expected to provide a good estimate of the process matrix. In this section we demonstrate that for our simulated Toffoli gate the compressed sensing method works well even for a much smaller number of configurations,  $m_{\text{conf}} \ll 576$ .

For the analysis we have simulated experimental data corresponding to a noisy Toffoli gate by adding truncated Gaussian noise with a small amplitude to each of  $M = 13824$  ideal measurement probabilities  $P_i^{\text{ideal}}$ . We assumed the set of experimental probabilities in Eq. (5) to be of the form  $P_i^{\text{exp}} = P_i^{\text{ideal}} + \Delta P_i$ , where  $\Delta P_i$  are random numbers sampled from the normal distribution with zero mean and a small standard deviation  $\sigma$ . By choosing different values of the standard deviation  $\sigma$  we can change the process fidelity of the simulated Toffoli gate: a smaller value of  $\sigma$  makes the process fidelity closer to 1. After adding the Gaussian noise  $\Delta P_i$  to the ideal probabilities  $P_i^{\text{ideal}}$ , we check whether the resulting simulated probabilities  $P_i^{\text{exp}}$  are in the interval  $[0, 1]$ . If a  $P_i^{\text{exp}}$  happens to be outside the interval  $[0, 1]$ , we repeat the procedure until the condition  $P_i^{\text{exp}} \in [0, 1]$  is satisfied. Finally, we renormalize each set of 8 probabilities corresponding to the same measurement configuration so that these probabilities add up to 1.

Thus the simulated imperfect quantum process is defined by  $M = 13824$  probabilities, corresponding to  $M_{\text{conf}} = 1728$  configurations; the process fidelity for a particular realization (used here) with  $\sigma = 0.01$  is  $F_{\chi} = F(\chi_{\text{full}}, \chi_{\text{ideal}}) = 0.959$ . We then test efficiency of the compressed sensing method by randomly selecting  $m_{\text{conf}} \leq 1728$  configurations, finding the corresponding process matrix  $\chi_{\text{CS}}$ , and comparing it with the full-data matrix  $\chi_{\text{full}}$  by calculating the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$ . We also calculate the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  given by  $\chi_{\text{CS}}$ .

The red line in Fig. 8 shows the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  as a function of the number  $m_{\text{conf}}$  of randomly selected configurations. The value of  $\varepsilon$  is chosen to be practically equal to  $\varepsilon_{\text{opt}} = \|(\vec{P}_{\text{full}}^{\text{exp}} - \Phi \vec{\chi}_{\text{full}})\|_{\ell_2} / \sqrt{M} = 0.01146$  (the relative difference is less than  $10^{-3}$ ). The  $\ell_1$ -minimization is done using the CVX-SeDuMi package. The error bars are calculated by repeating the procedure of random selection 7 times. We see a reasonably high fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  of the reconstructed process matrix even for small numbers of selected configurations. For example,  $F(\chi_{\text{CS}}, \chi_{\text{full}}) = 0.95$  for only  $m_{\text{conf}} = 40$  configurations, which represents a reduction by more than a factor of 40 compared with the full QPT and approximately a factor of 15 compared with the threshold of the underdetermined system of equations.

The blue line in Fig. 8 shows the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  calculated by the CS method. We see that it remains practically flat down to  $m_{\text{conf}} \gtrsim 40$ , which means

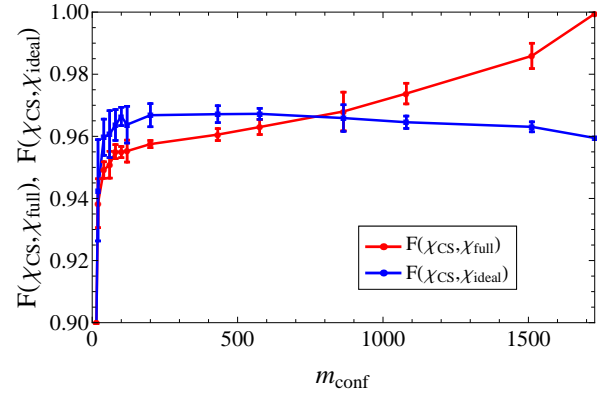


FIG. 8: (color online) CS QPT for a simulated Toffoli gate. Red line: fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  of the process matrix estimation, blue line: the estimated process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$ , both as functions of the data set size, expressed as the number  $m_{\text{conf}}$  of randomly selected configurations. The full QPT corresponds to 1728 configurations. The system of equations becomes underdetermined when  $m_{\text{conf}} < 576$ .

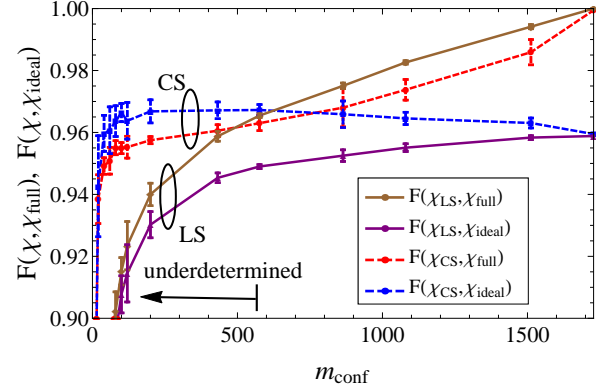


FIG. 9: (color online) Comparison between the calculations using CS and LS methods for the simulated Toffoli gate. Solid lines are for the LS method, dashed lines (the same as in Fig. 8) are for the CS method. In the underdetermined regime the CS-method results are much better than the LS-method results.

that  $\chi_{\text{CS}}$  can be used efficiently to estimate the actual process fidelity.

Figure 9 shows similar results calculated using the LS method (for comparison the lines from Fig. 8 are shown by dashed lines). We see that the LS method still works in the underdetermined regime ( $m_{\text{conf}} < 576$ ); however, it works significantly worse than the CS method. As an example, for  $m_{\text{conf}} = 40$  the fidelity of the process matrix estimation using the LS method is  $F(\chi_{\text{LS}}, \chi_{\text{full}}) = 0.86$ , which is significantly less than  $F(\chi_{\text{CS}}, \chi_{\text{full}}) = 0.95$  for the CS method. Similarly, for  $m_{\text{conf}} = 40$  the process fidelity obtained via the CS method,  $F(\chi_{\text{CS}}, \chi_{\text{ideal}}) = 0.96$  is close to the full-data value of 0.959, while the LS-method value,  $F(\chi_{\text{LS}}, \chi_{\text{ideal}}) = 0.85$ , is quite different.

Besides using the Pauli-error basis for the results shown in Fig. 8, we have also performed the calculations using the SVD

basis. The results (not shown) are very close to those in Fig. 8, and the relative fidelity  $F(\chi_{\text{CS-SVD}}, \chi_{\text{CS}})$  is above 0.98 for  $m_{\text{conf}} > 200$  and above 0.95 for  $m_{\text{conf}} > 40$ . We have also performed the calculations using non-optimal values of the noise parameter  $\varepsilon$ . In comparison with the results for CZ gate shown in Fig. 5, the results for the Toffoli gate (not shown) are more sensitive to the variation of  $\varepsilon$ . In particular, the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$  is about 0.93 for  $\varepsilon = 1.2\varepsilon_{\text{opt}}$  (not significantly depending on  $m_{\text{conf}}$  for  $m_{\text{conf}} > 40$ ) and the process fidelity  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  for  $\varepsilon = 1.2\varepsilon_{\text{opt}}$  is approximately 0.93 instead of the actual value 0.96.

Compared with the two-qubit case, it takes significantly more computing time and memory to solve the  $\ell_1$ -minimization problem for three qubits. In particular, our calculations in the Pauli-error basis took about 8 hours per point on a personal computer for  $m_{\text{conf}} \simeq 1500$  and about 1.5 hours per point for  $m_{\text{conf}} \simeq 40$ ; this is three orders of magnitude longer than for two qubits. The amount of used computer memory was 3–10 GB, which is two orders of magnitude larger than for two qubits. (The calculations in the SVD basis for the Toffoli gate took 1–3 hours per point and  $\sim 2$  GB of memory.) Such a strong scaling of required computer resources with the number of qubits seems to be the limiting factor in extending the CS QPT beyond three qubits, unless a more efficient algorithm is found. (Note that LS calculations required similar amount of memory, but the computation time was much shorter.)

The presented results have been obtained using the CVX-SeDuMi package. We also attempted to use the YALMIP-SDPT3 package. However, in our realization of computation the calculation results were very unreliable for  $m_{\text{conf}} < 200$  using the SVD basis, and even worse when the Pauli-error basis was used. Therefore we decided to use only the CVX-SeDuMi package for the 3-qubit CS procedure.

## VI. STANDARD DEVIATION OF STATE FIDELITY

As shown in previous sections, the process matrices  $\chi_{\text{CS}}$  obtained via the CS method allow us to estimate reliably the process fidelity  $F_\chi = F(\chi, \chi_{\text{ideal}})$  of a gate using just a small fraction of the full experimental data. While  $F_\chi$  is the most widely used characteristic of an experimental gate accuracy, it is not the only one. An equivalent characteristic (usually used in randomized benchmarking) is the average state fidelity, defined as  $\overline{F}_{\text{st}} = \int \text{Tr}(\rho_{\text{actual}} \rho_{\text{ideal}}) d|\psi_{\text{in}}\rangle / \int d|\psi_{\text{in}}\rangle$ , where the integration is over the initial pure states  $|\psi_{\text{in}}\rangle$  (using the Haar measure; it is often assumed that  $\int d|\psi_{\text{in}}\rangle = 1$ ), while the states  $\rho_{\text{ideal}}$  and  $\rho_{\text{actual}}$  are the ideal and actual final states for the initial state  $|\psi_{\text{in}}\rangle$ . The average state fidelity  $\overline{F}_{\text{st}}$  is sometimes called the “gate fidelity” [18] and can be naturally measured in the randomized benchmarking ( $F_{\text{RB}} = \overline{F}_{\text{st}}$ ); it is linearly related [68, 69] to the process fidelity,  $\overline{F}_{\text{st}} = (F_\chi d + 1)/(d + 1)$ , where  $d = 2^N$  is the Hilbert space dimension.

Besides the average state fidelity, an obviously important characteristic of a gate operation is the worst-case state fidelity  $F_{\text{st}, \text{min}}$ , which is minimized over the initial state. Un-

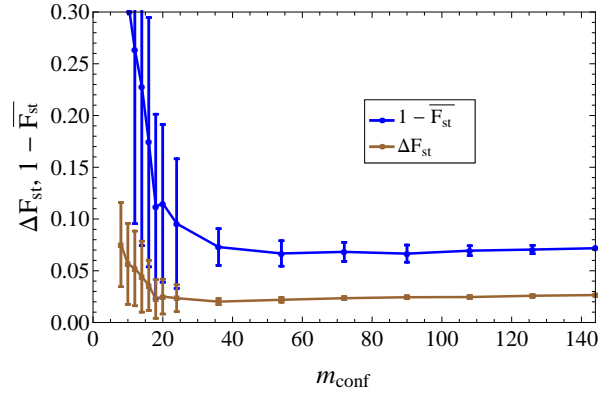


FIG. 10: (color online) Blue (upper) line: average state infidelity  $1 - \overline{F}_{\text{st}}$  for the CS-estimated process matrix  $\chi_{\text{CS}}$  as a function of the selected data set size for the experimental CZ gate (this line is linearly related to the blue line in Fig. 2). Brown (lower) line: the standard deviation of the state fidelity  $\Delta F_{\text{st}}$ , defined via variation of the initial state, Eq. (16), using the same  $\chi_{\text{CS}}$ . The error bars are computed by repeating the procedure 50 times with different random selections of used configurations.

fortunately, the minimum state fidelity is hard to find computationally even when the process matrix  $\chi$  is known. Another natural characteristic is the standard deviation of the state fidelity,

$$\Delta F_{\text{st}} = \sqrt{\overline{F_{\text{st}}^2} - \overline{F}_{\text{st}}^2}, \quad (16)$$

where  $\overline{F_{\text{st}}^2} = \int [\text{Tr}(\rho_{\text{actual}} \rho_{\text{ideal}})]^2 d|\psi_{\text{in}}\rangle / \int d|\psi_{\text{in}}\rangle$  is the average square of the state fidelity. The advantage of  $\Delta F_{\text{st}}$  in comparison with  $F_{\text{st}, \text{min}}$  is that  $\overline{F_{\text{st}}^2}$  and  $\Delta F_{\text{st}}$  can be calculated from  $\chi$  in a straightforward way [70, 71]. Our way of calculating  $\overline{F_{\text{st}}^2}$  is described in Appendix C [see Eq. (C10)].

We have analyzed numerically how well the CS QPT estimates  $\Delta F_{\text{st}}$  from the reduced data set, using the previously calculated process matrices  $\chi_{\text{CS}}$  for the experimental CZ gate and the simulated Toffoli gate (considered in Secs. IV and V). The results are presented in Figs. 10 and 11. We show the average state infidelity,  $1 - \overline{F}_{\text{st}}$ , and the standard deviation of the state fidelity,  $\Delta F_{\text{st}}$ , as functions of the number of selected configurations,  $m_{\text{conf}}$ . The random selection of used configurations is repeated 50 times for Fig. 10 (7 times for Fig. 11), the error bars show the statistical variation, while the dots show the average values.

As seen from Figs. 10 and 11, the CS method estimates reasonably well not only the average state fidelity  $\overline{F}_{\text{st}}$  (which is equivalent to  $F_\chi$  presented in Figs. 2 and 8), but also its standard deviation  $\Delta F_{\text{st}}$ . It is interesting to note that  $\Delta F_{\text{st}}$  is significantly smaller than the infidelity  $1 - \overline{F}_{\text{st}}$ , which means that the state fidelity  $\text{Tr}(\rho_{\text{actual}} \rho_{\text{ideal}})$  does not vary significantly for different initial states [the ratio  $\Delta F_{\text{st}}/(1 - \overline{F}_{\text{st}})$  is especially small for the simulated Toffoli gate, though this may be because of our particular way of simulation].



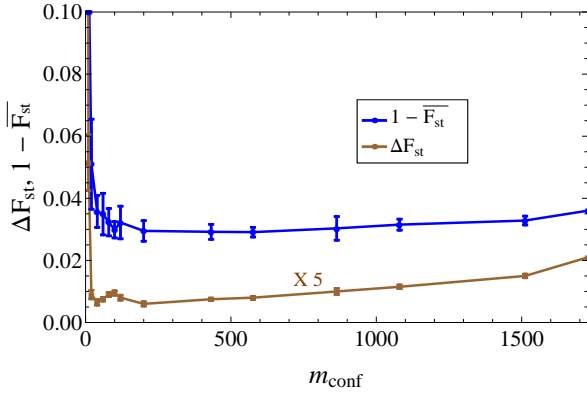


FIG. 11: (color online) The same as in Fig. 10, but for the simulated Toffoli gate. The random selection of configurations is repeated 7 times for each point. The results for the standard deviation  $\Delta F_{\text{st}}$  are multiplied by the factor of 5 for clarity.

## VII. CONCLUSION

In this paper we have numerically analyzed the efficiency of compressed sensing quantum process tomography (CS QPT) [24, 25] applied to superconducting qubits (we did not consider the CS method of Refs. [31, 32]). We have used experimental data for two-qubit CZ gates realized with Xmon and phase qubits, and simulated data for the three-qubit Toffoli gate with numerically added noise. We have shown that CS QPT permits a reasonably high fidelity estimation of the process matrix from a substantially reduced data set compared to the full QPT. In particular, for the CZ gate (Fig. 2) the amount of data can be reduced by a factor of  $\sim 7$  compared to the full QPT (which is a factor of  $\sim 4$  compared to the threshold of underdetermined system of equations). For the Toffoli gate (Fig. 8) the data reduction factor is  $\sim 40$  compared to the full QPT ( $\sim 15$  compared to the threshold of underdeterminacy).

In our analysis we have primarily used two characteristics. The first characteristic is the comparison between the CS-obtained process matrix  $\chi_{\text{CS}}$  and the matrix  $\chi_{\text{full}}$  obtained from the full data set; this comparison is quantitatively represented by the fidelity  $F(\chi_{\text{CS}}, \chi_{\text{full}})$ . The second characteristic is how well the CS method estimates the process fidelity  $F_{\chi}$ , i.e., how close  $F(\chi_{\text{CS}}, \chi_{\text{ideal}})$  is to the full-data value  $F(\chi_{\text{full}}, \chi_{\text{ideal}})$ . Besides these two characteristics, we have also calculated the standard deviation of the state fidelity  $\Delta F_{\text{st}}$  [Eq. (16)] and checked how well the CS method estimates  $\Delta F_{\text{st}}$  from a reduced data set (Figs. 10 and 11). Our compressed sensing method depends on the choice of the basis, in which the process matrix should be approximately sparse, and also depends on the choice of the noise parameter  $\varepsilon$  [see Eq. (11)]. We have used two bases: the Pauli-error basis and the SVD basis. The results obtained in both bases are similar to each other, though the SVD basis required less computational resources. The issue of choosing proper  $\varepsilon$  is not trivial. In our calculations we have used a value slightly larger than the noise level calculated from the full data set. However, in an experiment with a reduced data set this way of choosing  $\varepsilon$  is

not possible, so its value should be chosen from an estimate of the inaccuracy of the experimental probabilities. We have shown that the CS method tolerates some inaccuracy of  $\varepsilon$  (up to  $\sim 60\%$  for the results shown in Fig. 5); however, finding a proper way of choosing  $\varepsilon$  is still an open issue.

We have also compared the performance of the CS method with the least squares optimization. Somewhat surprisingly, the LS method can still be applied when the systems of equations (5) is underdetermined (unless the data set size is too small). This is because the condition of a process matrix being physical (positive, trace-preserving) usually makes satisfying Eqs. (5) impossible. However, even though the LS method formally works, it gives a less accurate estimate of  $\chi$  than the CS method in the significantly underdetermined regime (although it does give a better estimate in the overdetermined regime). The advantage of the CS method over the LS method is more pronounced for the Toffoli gate (Fig. 9).

Thus the CS QPT is useful for two-qubit and three-qubit quantum gates based on superconducting qubits. The method offers a very significant reduction of the needed amount of experimental data. However, the scaling of the required computing resources with the number of qubits seems to be prohibitive: in our calculations it took three orders of magnitude longer and two orders of magnitude more memory for the three-qubit-gate calculation than for two qubits. Such a scaling of computing resources seems to be a limiting factor in the application of our implementation of the CS method for QPT of four or more qubits. Therefore, the development of more efficient numerical algorithms for the CS QPT is an important task for future research.

## ACKNOWLEDGEMENTS

The authors thank Yuri Bogdanov, Steven Flammia, Justin Dressel, and Eyob Sete for useful discussions. We also thank Matteo Mariani for being involved in this work at its early stage. The research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office Grant No. W911NF-10-1-0334. All statements of fact, opinion, or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government. We also acknowledge support from the ARO MURI Grant No. W911NF-11-1-0268.

## Appendix A: Pauli-error basis

In this Appendix we discuss the definition of the Pauli-error basis used in this paper. The detailed theory of the QPT in the Pauli-error basis is presented in Ref. [8].

Let us start with description of a quantum process  $\mathcal{E}$  in the Pauli basis  $\{\mathcal{P}_{\alpha}\}$ ,

$$\rho^{\text{in}} \mapsto \mathcal{E}(\rho^{\text{in}}) = \sum_{\alpha, \beta=1}^{d^2} \chi_{\alpha\beta} \mathcal{P}_{\alpha} \rho^{\text{in}} \mathcal{P}_{\beta}^{\dagger}, \quad (\text{A1})$$

where for generality  $\mathcal{P}$  is not necessarily Hermitian (to include the modified Pauli basis, in which  $Y = -i\sigma_y$ ). Recall that  $d = 2^N$  is the dimension of the Hilbert space for  $N$  qubits.

In order to compare the process  $\mathcal{E}$  with a desired unitary rotation  $U$  [i.e. with the map  $\mathcal{U}(\rho^{\text{in}}) = U\rho^{\text{in}}U^\dagger$ ], let us formally apply the inverse unitary  $U^{-1} = U^\dagger$  after the process  $\mathcal{E}$ . The resulting composed process

$$\tilde{\mathcal{E}} = \mathcal{U}^{-1} \circ \mathcal{E} \quad (\text{A2})$$

characterizes the error: if  $\mathcal{E}$  is close to the desired  $\mathcal{U}$ , then  $\tilde{\mathcal{E}}$  is close to the identity (memory) operation. The process matrix  $\tilde{\chi}$  of  $\tilde{\mathcal{E}}$  in the Pauli basis is what we call in this paper the process matrix in the Pauli-error basis.

The process matrix  $\tilde{\chi}$  obviously satisfies the relation

$$\sum_{\alpha,\beta} \tilde{\chi}_{\alpha\beta} \mathcal{P}_\alpha \rho^{\text{in}} \mathcal{P}_\beta^\dagger = U^{-1} \left( \sum_{\alpha,\beta} \chi_{\alpha\beta} \mathcal{P}_\alpha \rho^{\text{in}} \mathcal{P}_\beta^\dagger \right) U, \quad (\text{A3})$$

which can be rewritten as

$$\sum_{\alpha,\beta} \tilde{\chi}_{\alpha\beta} (U \mathcal{P}_\alpha) \rho^{\text{in}} (U \mathcal{P}_\beta)^\dagger = \sum_{\alpha,\beta} \chi_{\alpha\beta} \mathcal{P}_\alpha \rho^{\text{in}} \mathcal{P}_\beta^\dagger. \quad (\text{A4})$$

Therefore the error matrix  $\tilde{\chi}$  is formally the process matrix of the original map  $\mathcal{E}$ , expressed in the operator basis

$$E_\alpha = U \mathcal{P}_\alpha. \quad (\text{A5})$$

This is the Pauli-error basis used in our paper. (Another obvious way to define the error basis is to use  $E_\alpha = \mathcal{P}_\alpha U$  [8]; however, we do not use this second definition here.) The Pauli-error basis matrices  $E_\alpha$  have the same normalization as the Pauli matrices,

$$\langle E_\alpha | E_\beta \rangle = \text{Tr}(E_\alpha^\dagger E_\beta) = d \delta_{\alpha\beta}. \quad (\text{A6})$$

The matrices  $\chi$  and  $\tilde{\chi}$  (in the Pauli and Pauli-error bases) are related via unitary transformation,

$$\tilde{\chi} = V \chi V^\dagger, \quad V_{\alpha\beta} = \text{Tr}(\mathcal{P}_\alpha^\dagger U^\dagger \mathcal{P}_\beta) / d. \quad (\text{A7})$$

The matrix  $\tilde{\chi}$  has a number of convenient properties [8]. It has only one large element, which is at the upper left corner and corresponds to the process fidelity,  $\tilde{\chi}_{\mathcal{I}\mathcal{I}} = F_\chi = F(\chi, \chi_{\text{ideal}})$ . All other non-zero elements of  $\tilde{\chi}$  describe imperfections. In particular, the imaginary elements in the left column (or upper row) characterize unitary imperfections (assuming the standard non-modified Pauli basis), other off-diagonal elements are due to decoherence, and the diagonal elements correspond to the error probabilities in the Pauli-twirling approximation.

## Appendix B: SVD basis

The SVD basis used in this paper is introduced following Ref. [24]. Let us start with the so-called natural basis for  $d \times d$  matrices, which consists of matrices  $E_\alpha^{\text{nat}}$ , having one element

equal to one, while other elements are zero. The numbering corresponds to the vectorized form obtained by stacking the columns: for  $\alpha = (d-1)i + j$  the matrix is  $(E_\alpha^{\text{nat}})_{lk} = \delta_{il} \delta_{jk}$ . For a desired unitary rotation  $U$ , the process matrix  $\chi^{\text{nat}}$  in the natural basis can be obtained by expanding  $U$  in the natural basis,  $U = \sum_\alpha u_\alpha E_\alpha^{\text{nat}}$ , and then constructing the outer product,

$$\chi_{\alpha\beta}^{\text{nat}} = u_\alpha u_\beta^*. \quad (\text{B1})$$

For example, for the ideal CZ gate the components  $u_\alpha$  are  $(1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, -1)$ , and  $\chi^{\text{nat}}$  has 16 non-zero elements, equal to  $\pm 1$ . Note that  $\chi^{\text{nat}}$  is a rank-1 matrix with  $\text{Tr}(\chi^{\text{nat}}) = \sum_\alpha |u_\alpha|^2 = d$ .

We then apply numerical procedure of the SVD decomposition, which diagonalizes the matrix  $\chi^{\text{nat}}$  for the desired unitary process,

$$\chi^{\text{nat}} = V \text{diag}(d, 0, \dots, 0) V^\dagger, \quad (\text{B2})$$

where  $V$  is a unitary  $d^2 \times d^2$  matrix and the only non-zero eigenvalue is equal to  $d$  because  $\text{Tr}(\chi^{\text{nat}}) = d$ . The columns of thus obtained transformation matrix  $V$  are the vectorized forms of thus introduced SVD-basis matrices  $E_\alpha^{\text{SVD}}$ ,

$$E_\alpha^{\text{SVD}} = \sum_{\beta=1}^{d^2} V_{\beta\alpha} E_\beta^{\text{nat}}. \quad (\text{B3})$$

Note that the notation  $V$  used in Appendix A has a different meaning.

The matrices of the SVD basis introduced via Eqs. (B2) and (B3) have the different normalization compared with the Pauli basis,

$$\text{Tr}(E_\alpha^{\text{SVD}\dagger} E_\beta^{\text{SVD}}) = \delta_{\alpha\beta}. \quad (\text{B4})$$

Correspondingly, the normalization of the process matrix  $\chi^{\text{SVD}}$  in the SVD basis is  $\text{Tr} \chi^{\text{SVD}} = d$  (for a trace-preserving process). For the ideal unitary process the matrix  $\chi^{\text{SVD}}$  has one non-zero (top left) element, which is equal to  $\sqrt{d}$ . For an imperfect realization of the desired unitary operation the top left element is related to the process fidelity as  $\chi_{11}^{\text{SVD}} = F_\chi$ .

Note that when the numerical SVD procedure (B2) is applied to  $\chi^{\text{nat}}$  of ideal CZ and/or Toffoli gates, many (most) of the resulting SVD-basis matrices  $E_\alpha^{\text{SVD}}$  coincide with the matrices of the natural basis  $E_\alpha^{\text{nat}}$ . Since these matrices contain only one non-zero element, the matrix  $\Phi$  in Eq. (5) is simpler (has more zero elements) than for the Pauli or Pauli-error basis. (The number of non-zero elements of  $\Phi$  in the SVD basis is crudely twice less for the CZ gate and 4 times less for the Toffoli gate.) As the result, from the computational point of view it is easier to use the SVD basis than the Pauli-error basis: less memory and less computational time are needed.

## Appendix C: Average square of state fidelity

In this subsection we present a detailed derivation of an explicit formula for the squared state fidelity  $\overline{F_{\text{st}}^2}$ , averaged over

all pure initial states, for a quantum operation, represented via Kraus operators. We follow the same steps as in Ref. [71], where a closed-form expression for  $\overline{F_{\text{st}}^2}$  in terms of the process matrix  $\chi$  was presented. Although our approach is not new, we show it here for completeness.

We begin by writing the quantum operation as  $\mathcal{E} = \mathcal{U} \circ \tilde{\mathcal{E}}$  [see Eq. (A2)], where  $\mathcal{U}$  corresponds to the ideal (desired) unitary operation, while the map  $\tilde{\mathcal{E}}$  accounts for the errors in the actual gate. Let

$$\tilde{\mathcal{E}}(\rho) = \sum_n A_n \rho A_n^\dagger \quad (\text{C1})$$

be the operator-sum representation of  $\tilde{\mathcal{E}}$ , where  $\{A_n\}_{n=1}^{d^2}$  are Kraus operators satisfying the trace-preservation condition  $\sum_n A_n^\dagger A_n = \mathbb{I}$ . The Kraus operators can be easily obtained from the process matrix  $\chi_{\alpha\beta}$  describing the operation  $\mathcal{E}$ . Note that by diagonalizing  $\chi$ , i.e.,  $\chi = V D V^\dagger$ , where  $V$  is unitary and  $D = \text{diag}(\lambda_1, \lambda_2, \dots)$  with  $\lambda_n \geq 0$ , we can express the Kraus operators in Eq. (C1) as  $A_n = \sqrt{\lambda_n} U^\dagger \sum_\alpha E_\alpha V_{\alpha n}$ , where  $U$  is the desired unitary.

Now, the state fidelity  $F_\phi$  (assuming a pure initial state  $|\phi\rangle$ ) can be written in terms of  $\{A_n\}$  as follows:

$$F_\phi \equiv \langle \phi | \tilde{\mathcal{E}}(\phi) | \phi \rangle = \sum_n \langle \phi | A_n | \phi \rangle \langle \phi | A_n^\dagger | \phi \rangle. \quad (\text{C2})$$

Notice that by using the identity  $\text{Tr}(A \otimes B) = \text{Tr}(A) \text{Tr}(B)$ , one can rewrite the above expression for  $F_\phi$  as

$$F_\phi = \sum_n \text{Tr}[(A_n \otimes A_n^\dagger)(|\phi\rangle \langle \phi|^{\otimes 2})], \quad (\text{C3})$$

where the notation  $|\phi\rangle \langle \phi|^{\otimes k} \equiv \underbrace{|\phi\rangle \langle \phi| \otimes |\phi\rangle \langle \phi| \otimes \dots \otimes |\phi\rangle \langle \phi|}_k$

means that the state is copied in  $k$  identical Hilbert spaces. Similarly, one can express the squared state fidelity as

$$\begin{aligned} F_\phi^2 &= \sum_{n,m} \langle \phi | A_n | \phi \rangle \langle \phi | A_n^\dagger | \phi \rangle \langle \phi | A_m | \phi \rangle \langle \phi | A_m^\dagger | \phi \rangle \\ &= \sum_{n,m} \text{Tr}[(A_n \otimes A_n^\dagger \otimes A_m \otimes A_m^\dagger)(|\phi\rangle \langle \phi|^{\otimes 4})]. \end{aligned} \quad (\text{C4})$$

In order to compute the average state fidelity  $\overline{F_{\text{st}}} = \int F_\phi d\phi$ , the average square of the state fidelity  $\overline{F_{\text{st}}^2} = \int F_\phi^2 d\phi$ , and higher powers of  $F_{\text{st}}$  (we assume the normalized integration over the initial pure states,  $\int d\phi = 1$ ), one can use the following result [72]

$$\int |\phi\rangle \langle \phi|^{\otimes k} d\phi = \frac{1}{\binom{k+d-1}{d-1}} \Pi_k, \quad \Pi_k \equiv \frac{1}{k!} \sum_{\sigma \in S_k} P_\sigma. \quad (\text{C5})$$

Here  $\sigma$  is an element of the permutation group  $S_k$  (the  $k!$  permutations of  $k$  objects) and the operator  $P_\sigma$  is the representation of  $\sigma$  in  $\mathcal{H}^{\otimes k} = \underbrace{\mathcal{H} \otimes \dots \otimes \mathcal{H}}_k$ , i.e.,

$$P_\sigma(|\phi_1\rangle \otimes |\phi_2\rangle \dots \otimes |\phi_k\rangle) = |\phi_{\sigma(1)}\rangle \otimes |\phi_{\sigma(2)}\rangle \dots \otimes |\phi_{\sigma(k)}\rangle. \quad (\text{C6})$$

(The operator  $P_\sigma$  acts on the wavefunction of  $kN$  qubits by permuting  $k$  blocks, each containing  $N$  qubits.)

In view of the above discussion, we see that the  $k$ th moment  $\overline{F_{\text{st}}^k} \equiv \int F_\phi^k d\phi$  can be expressed as a sum of  $(2k)!$  terms corresponding to the elements in  $S_{2k}$  [note that  $k$  in Eqs. (C5) and (C6) is now replaced with  $2k$ ],

$$\overline{F_{\text{st}}^k} = \frac{\sum_{n_1 \dots n_k} \sum_{\sigma \in S_{2k}} \text{Tr}[(A_{n_1} \otimes A_{n_1}^\dagger \otimes \dots \otimes A_{n_k} \otimes A_{n_k}^\dagger) P_\sigma]}{\binom{2k+d-1}{d-1} (2k)!}. \quad (\text{C7})$$

For example, the average state fidelity  $\overline{F_{\text{st}}}$  is determined by the sum over  $S_2$ ,

$$\begin{aligned} \text{Tr}(A_n \otimes A_n^\dagger \Pi_2) &= \frac{1}{2} \sum_{\sigma \in S_2} \text{Tr}(A_n \otimes A_n^\dagger P_\sigma) \\ &= \frac{1}{2} \sum_{\sigma \in S_2} \sum_{i_1, i_2} \langle i_1, i_2 | A_n \otimes A_n^\dagger | \sigma(i_1), \sigma(i_2) \rangle \\ &= \frac{1}{2} \underbrace{(\text{Tr}(A_n) \text{Tr}(A_n^\dagger))}_{\text{identity}} + \underbrace{\text{Tr}(A_n A_n^\dagger)}_{\text{transposition}}, \end{aligned} \quad (\text{C8})$$

which yields the well-known result [69]

$$\overline{F_{\text{st}}} = \frac{1}{d(d+1)} \left( \sum_n |\text{Tr}(A_n)|^2 + d \right). \quad (\text{C9})$$

In order to express  $\overline{F_{\text{st}}^2}$  in terms of Kraus operators, it is convenient to write each element of the group  $S_4$  as a product of disjoint cycles. The 24 elements of the permutation groups  $S_4$  can be grouped as follows (we use the so-called cycle notation for permutations):

- Identity (1 element): (1)(2)(3)(4) (this notation means that no change of position occurs for all numbers in the sequence 1234);
- Transpositions (6 elements): (12), (13), (14), (23), (24), and (34) (this notations means that only two specified numbers in the sequence are exchanged);
- 3-cycles (8 elements): (123), (132), (124), (142), (134), (143), (234), and (243) [here the notation (123) means the permutation  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ , while the remaining number does not change];
- Products of transpositions (3 elements): (12)(34), (13)(24), and (14)(23) (two pairs of numbers exchange);
- 4-cycles (6 elements): (1234), (1243), (1324), (1342), (1423), and (1432) [here (1234) means the permutation  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ ].

This classification simplifies keeping track of the terms  $N_\sigma \equiv \sum_{n,m} \text{Tr}[(A_n \otimes A_n^\dagger \otimes A_m \otimes A_m^\dagger) P_\sigma]$  in Eq. (C7). The corresponding contributions to the sum  $\sum_{\sigma \in S_4} N_\sigma$  are



the following:

Identity:

$$\left(\sum_n |\text{Tr}(A_n)|^2\right)^2.$$

Transpositions:

$$2d \sum_n |\text{Tr}(A_n)|^2 + 2 \sum_{n,m} \text{Tr}(A_n A_m^\dagger) \text{Tr}(A_n^\dagger) \text{Tr}(A_m) \\ + \sum_{n,m} (\text{Tr}(A_n A_m) \text{Tr}(A_n^\dagger) \text{Tr}(A_m^\dagger) + h.c.).$$

3-cycles:

$$4 \sum_n |\text{Tr}(A_n)|^2 + 2 \sum_{n,m} (\text{Tr}(A_n A_m^\dagger A_m) \text{Tr}(A_n^\dagger) + h.c.).$$

Products of transpositions:

$$d^2 + \sum_{n,m} (|\text{Tr}(A_n A_m)|^2 + |\text{Tr}(A_n A_m^\dagger)|^2).$$

4-cycles:

$$3d + \sum_{n,m} \text{Tr}(A_n A_m^\dagger A_m A_n^\dagger) + 2 \sum_{n,m} \text{Tr}(A_n A_m A_n^\dagger A_m^\dagger).$$

(We used the trace-preservation condition  $\sum_n A_n^\dagger A_n = \mathbb{I}$ ). Substituting the above terms in Eq. (C7) (with  $k = 2$ ), we finally obtain the average square of the state fidelity,

$$\overline{F_{\text{st}}^2} = \frac{1}{d(d+1)(d+2)(d+3)} \left( d^2 + 3d \right. \\ + 2(d+2) \sum_n |\text{Tr}(A_n)|^2 + \left( \sum_n |\text{Tr}(A_n)|^2 \right)^2 \\ + \sum_{n,m} (|\text{Tr}(A_n A_m)|^2 + |\text{Tr}(A_n A_m^\dagger)|^2) \\ + 2 \sum_{n,m} \text{Tr}(A_n A_m A_n^\dagger A_m^\dagger) + \sum_{n,m} \text{Tr}(A_n A_n^\dagger A_m A_m^\dagger) \\ + 2 \sum_{n,m} \text{Tr}(A_n A_m^\dagger) \text{Tr}(A_n^\dagger) \text{Tr}(A_m) \\ + 2 \sum_{n,m} \text{Re}[\text{Tr}(A_n A_m) \text{Tr}(A_n^\dagger) \text{Tr}(A_m^\dagger)] \\ \left. + 4 \sum_{n,m} \text{Re}[\text{Tr}(A_n A_m^\dagger A_m A_n^\dagger) \text{Tr}(A_m)] \right). \quad (\text{C10})$$

This is the formula we used in this paper to calculate  $\overline{F_{\text{st}}^2}$ .

- 
- [1] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, Cambridge, England, 2000).
- [2] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455 (1997).
- [3] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390 (1997).
- [4] N. Boulant, T. F. Havel, M. A. Pravia, and D. G. Cory, Phys. Rev. A **67**, 042322 (2003).
- [5] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).
- [6] M. Mohseni and A. T. Rezakhani, Phys. Rev. A **80**, 010101 (2009).
- [7] A. G. Kofman and A. N. Korotkov, Phys. Rev. A **80**, 042103 (2009).
- [8] A. N. Korotkov, arXiv:1309.6405.
- [9] D. W. Leung, J. Math. Phys. **44**, 528 (2003).
- [10] G. M. D'Ariano and P. Lo Presti, Phys. Rev. Lett. **91**, 047902 (2003).
- [11] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laamme, Science **317**, 1893 (2007).
- [12] M. Mohseni and D. A. Lidar, Phys. Rev. Lett. **97**, 170501 (2006).
- [13] M. M. Wolf, J. Eisert, T. S. Cubitt, and J. I. Cirac, Phys. Rev. Lett. **101**, 150402 (2008).
- [14] Yu. I. Bogdanov, G. Brida, M. Genovese, S. P. Kulik, E. V. Moreva, and A. P. Shurupov, Phys. Rev. Lett. **105**, 010404 (2010).
- [15] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).
- [16] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Phys. Rev. A **77**, 012307 (2008).
- [17] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B **7**, S347 (2005).
- [18] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 080505 (2012).
- [19] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).
- [20] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).
- [21] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009).
- [22] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, Phys. Rev. A **87**, 030301(R) (2013).
- [23] L. Steffen, M. P. da Silva, A. Fedorov, M. Baur, and A. Wallraff, Phys. Rev. Lett. **108**, 260506 (2012).
- [24] R. Kosut, arXiv:0812.4323.
- [25] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Phys. Rev. Lett. **106**, 100401 (2011).
- [26] E. J. Candes, J. K. Romberg, T. Tao, Comm. Pure Appl. Math. **59**, 1207 (2006).
- [27] D. L. Donoho, IEEE Trans. Inf. Theory **52**, 1289 (2006).
- [28] E. J. Candes, C.R. Seances Acad. Sci. Ser. 1, **346**, 589 (2008).
- [29] E. J. Candes, M. B. Wakin, IEEE Signal Proc. Mag. **25**, 21 (2008).
- [30] Roughly speaking, a matrix is  $s$ -compressible if it can be well approximated by an  $s$ -sparse matrix.
- [31] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker and J. Eisert, Phys. Rev. Lett. **105**, 150401 (2010).
- [32] S. T. Flammia, D. Gross, Y.-K. Liu and J. Eisert, New J. Phys. **14**, 095022 (2012).
- [33] C. H. Baldwin, A. Kalev, and I. H. Deutsch, arXiv:1404.2877.
- [34] W.-T. Liu, T. Zhang, J.-Y. Liu, P.-X. Chen, and J.-M. Yuan,

- Phys. Rev. Lett. **108**, 170403 (2012).
- [35] A. Smith, C. A. Riofrio, B. E. Anderson, H. Sosa-Martinez, I. H. Deutsch, and P. S. Jessen, Phys. Rev. A **87**, 030102(R) (2013).
- [36] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
- [37] R. Kosut, I. A. Walmsley, H. Rabitz, arXiv:quant-ph/0411093.
- [38] M. Ježek, J. Fiurášek, and Z. Hradil, Phys. Rev. A **68**, 012305 (2003).
- [39] M. Riebe, K. Kim, P. Schindler, T. Monz, P. O. Schmidt, T. K. Körber, W. Hänsel, H. Häffner, C. F. Roos, and R. Blatt, Phys. Rev. Lett. **97**, 220407 (2006).
- [40] M. Mičuda, M. Sedláč, I. Straka, M. Miková, M. Dušek, M. Ježek, and J. Fiurášek, Phys. Rev. A **89**, 042304 (2014).
- [41] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
- [42] Yu. I. Bogdanov, M. V. Chekhova, L. A. Krivitsky, S. P. Kulik, A. N. Penin, A. A. Zhukov, L. C. Kwek, C. H. Oh, and M. K. Tey, Phys. Rev. A **70**, 042303 (2004).
- [43] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, Phys. Rev. Lett. **93**, 080502 (2004).
- [44] G. D. Fuchs, A. L. Falk, V. V. Dobrovitski, and D. D. Awschalom, Phys. Rev. Lett. **108**, 157602 (2012).
- [45] For a vector  $x \in C^m$ , the  $\ell_1$  and  $\ell_2$  norms are defined as  $\|x\|_{\ell_2} = \sqrt{x^\dagger x} = \sqrt{\sum_{i=1}^m |x_i|^2}$  and  $\|x\|_{\ell_1} = \sum_{i=1}^m |x_i|$ .
- [46] M. Mariani, H. Wang, T. Yamamoto, M. Neeley, R. C. Bialczak, Y. Chen, M. Lenander, E. Lucero, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, Y. Yin, J. Zhao, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Science **334**, 61 (2011).
- [47] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, M. Steffen, A. N. Cleland and J. M. Martinis, Nature Phys. **6**, 409 (2010).
- [48] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frinzio, S. M. Girvin and R. J. Schoelkopf, Nature **482**, 382 (2012).
- [49] A. Dewes, F. R. Ong, V. Schmitt, R. Lauro, N. Boulant, P. Bertet, D. Vion, and D. Esteve, Phys. Rev. Lett. **108**, 057002 (2012).
- [50] A. Fedorov, L. Steffen, M. Baur, M. P. da Silva, and A. Wallraff, Nature **481**, 170 (2012).
- [51] J. M. Chow, J. M. Gambetta, A. D. Córcoles, S. T. Merkel, J. A. Smolin, C. Rigetti, S. Poletto, G. A. Keefe, M. B. Rothwell, J. R. Rozen, M. B. Ketchen, and M. Steffen, Phys. Rev. Lett. **109**, 060501 (2012).
- [52] J. M. Chow, J. M. Gambetta, A. W. Cross, S. T. Merkel, C. Rigetti and M. Steffen, New J. Phys. **15**, 115012 (2013).
- [53] T. Yamamoto, M. Neeley, E. Lucero, R. C. Bialczak, J. Kelly, M. Lenander, M. Mariani, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, Y. Yin, A. N. Cleland, and J. M. Martinis, Phys. Rev. B **82**, 184515 (2010).
- [54] R. Barends, J. Kelly, A. Megrant, D. Sank, E. Jeffrey, Y. Chen, Y. Yin, B. Chiaro, J. Mutus, C. Neill, P. O'Malley, P. Roushan, J. Wenner, T. C. White, A. N. Cleland, and John M. Martinis, Phys. Rev. Lett. **111**, 080502 (2013).
- [55] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsforth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and John M. Martinis, Nature **508**, 500 (2014).
- [56] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Phys. Rev. Lett. **91**, 167005 (2003).
- [57] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsforth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Phys. Rev. Lett. **112**, 240504 (2014).
- [58] A. Uhlmann, Rep. Math. Phys. **9**, 273 - 279 (1976).
- [59] R. Jozsa, J. Mod. Opt. **41**, 2315 - 2323 (1994).
- [60] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).
- [61] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
- [62] M. C. Grant and S. P. Boyd, "CVX: Matlab software for disciplined convex programming", <http://cvxr.com/cvx>.
- [63] J. F. Sturm, Optimization Meth. & Soft., **11 & 12**, 625 (1999); the solver SeDuMi is available at <http://sedumi.ie.lehigh.edu>.
- [64] J. Löfberg, in Proceedings of the IEEE CACSD Symposium, p. 284 (Taipei, Taiwan, 2004); the package YALMIP is available at <http://users.isy.liu.se/johanl/yalmip>.
- [65] K. C. Toh, M. J. Todd, and R. H. Tütüncü, Optimization Meth. & Soft., **11 & 12**, 545 (1999); the solver SDPT3 is available at <http://www.math.nus.edu.sg/~matttohkc/sdpt3.html>.
- [66] D. G. Cory, M. D. Price, W. Maas, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo, Phys. Rev. Lett. **81**, 2152 (1998).
- [67] T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, Phys. Rev. Lett. **102**, 040501 (2009).
- [68] M. Horodecki, P. Horodecki, R. Horodecki, Phys. Rev. A **60**, 1888 (1999).
- [69] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).
- [70] L. H. Pedersen, N. M. Møller and K. Mølmer, Phys. Lett. A **372**, 7028 (2008).
- [71] E. Magesan, R. Blume-Kohout and J. Emerson, Phys. Rev. A **84**, 012309 (2011).
- [72] D. Poulin, R. Blume-Kohout, R. Laflamme, and H. Ollivier, Phys. Rev. Lett. **92**, 177906 (2004).